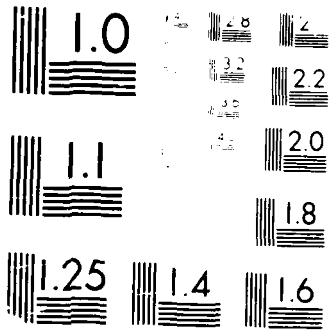
AD-A191 343 LEAVE-K-OUT DIAGNOSTICS FOR TIME SERIES(U) WASHINGTON UNIV SEATTLE DEPT OF STATISTICS A G BRUCE ET AL. SEP 87 TR-107 N00014-84-C-0169 1/1

UNCLASSIFIED F/G 12/3 NL



MICROCOPY RESOLUTION TEST CHART

4

LEAVE-K-OUT DIAGNOSTICS FOR TIME SERIES

by

Andrew G. Bruce and R. Douglas Martin

TECHNICAL REPORT No. 107

September 1987

Department of Statistics, GN-22

University of Washington

Seattle, Washington 98195 USA

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 107	2. GOVT ACCESSION NO. [illegible]
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Leave-K-Out Diagnostics for Time Series	5. TYPE OF REPORT & PERIOD COVERED: TR 12/1/83 - 5/31/88
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Andrew G. Bruce and R. Douglas Martin	8. CONTRACT OR GRANT NUMBER(s): N00014-84-C-0169
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, GN-22, University of Washington, Seattle, WA 98195	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR-661-003
11. CONTROLLING OFFICE NAME AND ADDRESS: ONR Code N63374, 1107 NE 45th Street, Seattle, WA 98105	12. REPORT DATE: September 1987
	13. NUMBER OF PAGES: 91
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We propose diagnostics for ARIMA model fitting for time series formed by deleting observations from the data and measuring the change in the estimates of the parameters. The use of leave-one-out diagnostics is a well establishe tool in regression analysis. We demonstrate the efficacy of observation deletion based diagnostics for ARIMA models, addressing issues special to the time diagnostics based on the innovations variance. It is shown that the dependency aspect of time series data gives rise to a /CONTINUED .....

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (continued)

"smearing" effect, which confounds the diagnostics for the coefficients. It is also shown that the diagnostics based on the innovations variance are much clearer and more sensitive than those for the coefficients. A "leave-k-out" diagnostics approach is proposed to deal with patches of outliers, and problems caused by "masking" are handled by use of iterative deletion. An overall strategy for ARIMA model fitting is given, and applied to two data sets.

LEAVE-K-OUT DIAGNOSTICS FOR TIME SERIES

Andrew G. Bruce
R. Douglas Martin

Department of Statistics, GN-22
University of Washington
Seattle, WA 98195

ABSTRACT

We propose diagnostics for ARIMA model fitting for time series formed by deleting observations from the data and measuring the change in the estimates of the parameters. The use of leave-one-out diagnostics is a well established tool in regression analysis. We demonstrate the efficacy of observation deletion based diagnostics for ARIMA models, addressing issues special to the time diagnostics based on the innovations variance. It is shown that the dependency aspect of time series data gives rise to a "smearing" effect, which confounds the diagnostics for the coefficients. It is also shown that the diagnostics based on the innovations variance are much clearer and more sensitive than those for the coefficients. A "leave-k-out" diagnostics approach is proposed to deal with patches of outliers, and problems caused by "masking" are handled by use of iterative deletion. An overall strategy for ARIMA model fitting is given, and applied to two data sets.

Research supported by NASC Contract N00014-86-K-0819, APL Contract N00014-84-K-0599, and ONR Contract N00014-84-C-0169.

1. Introduction

Regression diagnostics are becoming a well-accepted tool in the practice of statistics. This is evidenced not only by books devoted to the subject (e.g., Belsley, Kuh and Welsch, 1980; Cook and Weisberg, 1982; Atkinson, 1985), but also by the penetration of the concepts into standard texts on regression (e.g., Weisberg, 1980) and the increasingly widespread availability of software for computing the diagnostics. One also sees the basic leave-one-out diagnostic idea for linear regression begin carried over to somewhat more complicated settings such as logistic regression (Pregibon, 1981) and Cox regression (Storer and Crowley, 1985).

However, the literature appears to be relatively devoid of analogous results in the time-series setting, in spite of a rather obvious way to obtain leave-one-out diagnostics in the context of ARIMA model fitting for time series, at least in principle: One deletes a single observation at a time, and for each deletion computes a Gaussian maximum likelihood estimate for missing data (see, for example, Jones, 1980; Harvey and Pierse, 1984; Kohn and Ansley, 1986). It should be noted that use of Gaussian MLE's for missing data entails intuitively appealing use of predictions in place of missing data. A diagnostic display is obtained by comparing the leave-one-out MLE's with the Gaussian MLE's for the full data set versus time, on an appropriate comparison scale. This idea was articulated some time ago by Brillinger (1966), but only the advent of powerful computers and algorithms for fitting ARMA and ARIMA models with missing data has placed actual use of the procedure within reach.

In this paper we demonstrate the efficacy of observation deletion diagnostics for time series, addressing in the process some issues which are special to the time series setting. In particular, we consider not only diagnostics based on ARIMA model coefficients, but also diagnostics based on the innovations variance. We show that the time series problem gives rise to a "smearing" effect which is not encountered in the usual independent-observation setting. For diagnostics based on coefficients, this smearing can result in considerable

ambiguity concerning the numbers and locations of outliers. By both examples and by theoretical calculations, we show that diagnostics based on the innovations variance is far superior to coefficient-based diagnostics in this regard.

Furthermore outliers frequently occur in patches in the time series setting. Thus we proposed a "leave-k-out" diagnostic approach which is both effective and within computational reach.

The paper is organized as follows. Gaussian maximum likelihood estimation of ARIMA models with missing data is reviewed in Section 2. Section 3 presents the basic "leave-k-out" diagnostic, based on the coefficients and innovations variance, including a proposal for scaling. Some artificial examples are given which illustrate that the innovations variance is a better diagnostic tool. Analytical results on "smearing" effects associated with leave-k-out diagnostics are presented in Section 4. The problem of outlier type identification is discussed briefly in Section 5. Section 6 presents an iterative deletion procedure to overcome problems caused by masking. Techniques are also discussed for handling other types of disturbances, such as level shifts and variance changes. Finally, we give an overall strategy for ARIMA model identification and fitting using the leave-k-out diagnostics. This strategy is applied in Section 7 to two real data sets. Finally, possible extensions, computational aspects, scaling issues, and a connection with robust filtering are briefly mentioned in Section 8.

2. Estimation of ARIMA Models with Missing Data

Exact maximum likelihood estimates with missing data can be obtained using the state space representation of an ARIMA model. Various formulations have been given by Jones (1980), Harvey and Pierse (1984) and Kohn and Ansley (1986). The Harvey and Pierse approach is used here. The Kohn and Ansley approach has an attractive feature which we comment on in Section 8.

2.1 The Model

Consider a nonstationary process x_t, $t=1,\ldots,n$, which can be represented by an ARIMA $(p,d,q)\times(P,D,Q)$ model

$$\Phi(B^s)\phi(B)\nabla^d\nabla_s^D x_t = \gamma+\Theta(B^s)\theta(B)\varepsilon_t \tag{2.1}$$

where the ε_t, are the innovations. These are assumed to be independent normal random variables with zero mean and variance σ^2. B is the backshift operator, and the regular and seasonal difference operators are $\nabla=(1-B)$, $\nabla_s=(1-B^s)$, respectively. The intercept term is γ, the ordinary autoregressive and moving average operators are

$$\phi(B) = 1-\phi_1 B-\cdots-\phi_p B^p, \quad \theta(B) = 1-\theta_1 B-\cdots-\theta_q B^q \tag{2.2}$$

and the corresponding "seasonal" operators are

$$\Phi(B^s) = 1-\Phi_1 B^s-\cdots-\Phi_P B^{sP}, \quad \Theta(B^s) = 1-\Theta_1 B^s-\cdots-\Theta_Q B^{sQ}. \tag{2.2'}$$

Let α denote the $r\times 1$ vector of parameters,

$$\alpha' = (\phi_1,\cdots,\phi_p,\Phi_1,\cdots,\Phi_p,\theta_1,\cdots,\theta_q,\Theta_1,\cdots,\Theta_q)' \tag{2.3}$$

where $r=p+P+q+Q$, Assume that the polynomials in (2.2)–(2.2') have their roots outside the unit circle, so that the process $w_t\equiv\nabla^d\nabla_s^D x_t$ is stationary and invertible.

2.2 Kalman Filter Representation of the Likelihood Function

The state space formulation of (2.1) is based on the vector Markov state transition equation

$$\mathbf{x}_t = \mathbf{T}\mathbf{x}_{t-1} + \mathbf{r}\varepsilon_t \tag{2.4}$$

where $\mathbf{x}_t$ is an $m \times 1$ state vector, $\mathbf{T}$ is an $m \times m$ transition matrix, $\mathbf{r}$ is an $m \times 1$ vector, the ε_t are as in (2.1), and $m = \max(p + sP + d + sD, q + sQ + 1)$. The values of the process x_t are related to the state vector $\mathbf{x}_t$ via the noise-free observations equation

$$x_t = \mathbf{z}'\mathbf{x}_t \tag{2.5}$$

where $\mathbf{z}^T = (1, 0, \cdots, 0)$.

Let $\hat{\mathbf{x}}_t$ denote the optimal linear (Kalman) filter estimate of $\mathbf{x}_t$ (i.e., $\hat{\mathbf{x}}_t$ minimizes the mean-squared error and depends on the data $x_1, \cdots, x_t$), and let $\sigma^2 \mathbf{P}_t$ be the covariance matrix of filtering error $\hat{\mathbf{x}}_t - \mathbf{x}_t$. Also, let $\hat{\mathbf{x}}_{t|t-1} = \mathbf{T}\hat{\mathbf{x}}_{t-1}$ be the optimal one-step-ahead predictor of $\mathbf{x}_t$, and let $\sigma^2 \mathbf{P}_{t|t-1}$ denote the corresponding prediction error covariance matrix. The Kalman filter provides a well-known method for recursively evaluating $\hat{\mathbf{x}}_t$, $\hat{\mathbf{x}}_{t|t-1}$, $\mathbf{P}_t$, and $\mathbf{P}_{t|t-1}$:

$$\begin{aligned}
\hat{\mathbf{x}}_t &= \hat{\mathbf{x}}_{t|t-1} + f_t^{-1}\mathbf{P}_{t|t-1}\mathbf{z} \cdot e_t \\
\mathbf{P}_{t|t-1} &= \mathbf{T}\mathbf{P}_{t-1}\mathbf{T}' + \mathbf{r}\mathbf{r}'\sigma^2 \\
\mathbf{P}_t &= \mathbf{P}_{t|t-1} - f_t^{-1}\mathbf{P}_{t|t-1}\mathbf{z}\mathbf{z}'\mathbf{P}_{t|t-1}
\end{aligned} \tag{2.6}$$

where e_t is the observation-prediction residual

$$e_t \equiv x_t - E[x_t | x_1, \ldots, x_{t-1}] = x_t - \mathbf{z}'\hat{\mathbf{x}}_{t|t-1}$$

and f_t is the associated observation-prediction error variance

$$f_t \equiv E[e_t^2 | x_1, \ldots, x_{t-1}] = \mathbf{z}'\mathbf{P}_{t|t-1}\mathbf{z} = (\mathbf{P}_{t|t-1})_{11} . \tag{2.7}$$

The choice of initial condition is discussed shortly.

The log-likelihood is conveniently expressed in terms of e_t and f_t (see, for example, Harvey (1981)):

$$\log L(\mathbf{x}_n, \alpha, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{1}{2}\sum_{t=1}^{n}\log f_t - \frac{1}{2\sigma^2}\sum_{t=1}^{n} e_t^{\,2}/f_t \qquad (2.8)$$

where $\mathbf{x}_n = (x_1, \cdots, x_n)^T$. Note that $f_t = f_t(\alpha)$ and $e_t = e_t(\alpha)$.

Concentrating out σ^2 in (2.8), the maximum likelihood estimate $\hat{\alpha}$ is given by

$$\hat{\alpha} = argmin\{ \sum_{t=1}^{n} \log f_t(\alpha) + \log(\sum_{t=1}^{n} e_t^{\,2}(\alpha)/f_t(\alpha)) \} . \qquad (2.9)$$

If observation t_0 is missing, then the corresponding term in (2.9) is dropped, no update is performed in the Kalman filter, and $\hat{\mathbf{x}}_{t_0} = \hat{\mathbf{x}}_{t_0|t_0-1}$.

When a nonstationary ARIMA process is differenced to produce stationarity, the log-likelihood is given by (2.8) applied to the differenced series $w_t = \nabla^d \nabla_s^D x_t$.

2.3 The Special Case of Stationarity

Consider the special stationary ARMA process case of (2.1). Ignoring seasonal components, we use the state space formulation preferred by Harvey and Phillips (1979), among others, and choose $\mathbf{T}$ and $\mathbf{r}$ to be

$$\mathbf{T} = \left[\begin{array}{c|c} \phi_1 & \\ \cdots & \mathbf{I}_{m-1} \\ & \\ \phi_m & \mathbf{0}_{m-1}^T \end{array} \right]$$

$$\mathbf{r}' = (1, -\theta_1, \ldots, -\theta_{m-1})^T \qquad (2.10)$$

where $m = max(p, q+1)$, $\phi_i = 0$ for $i > p$ and $\theta_i = 0$ for $i > q+1$. Under stationarity

the initial conditions for (2.6) are $\hat{\mathbf{x}}_{1|0}=0$ and $\mathbf{P}_{1|0}$ which satisfies

$$\mathbf{P}_{1|0} = \mathbf{T}\mathbf{P}_{1|0}\mathbf{T}' + \mathbf{r}\mathbf{r}' \tag{2.11}$$

For numerical solution of (2.11), see Gardner, Harvey and Phillips (1980).

To ensure stationarity, the optimization of (2.9) over $\phi_1, \cdots, \phi_p$ must be constrained so that the roots of the polynomial equation $\phi(B)=0$ lie outside the unit circle. This is easily done by first reparameterizing in terms of the partial autoregressive coefficients, $b_i, i=1, \ldots, p$, and carrying out an unconstrained minimization over the transformed partial autogression coefficients u_i, $i = 1, \ldots, p$, where

$$b_i = \frac{1-e^{-u_i}}{1+e^{-u_i}}. \tag{2.12}$$

The parameters $\hat{\phi}_1, \cdots, \hat{\phi}_p$ are obtained by the Levinson (1947)–Durbin (1960) recursions. See Jones (1980) who also pointed out that invertibility of the estimated model can be assured by using partial moving average coefficients.

2.4 Nonstationary Models

Now consider the general ARIMA model without seasonal terms. One possible approach is to apply the state space model (2.10) to the differenced series w_t. However, with missing observations, this procedure is undesirable since the series w_t will have $(d+1)(D+1)$ times as many missing values as x_t. An alternative method due to Harvey and Pierse (1984) avoids this difficulty by utilizing a *levels* formulation of the state space model.

The levels formulation is based on separating the stationary and nonstationary parts of x_t in the state vector $\mathbf{x}_t$. Let $d_0 = d + sD$, $m^{(w)} = max(p, q+1)$, and $m = m^{(w)} + d_0$. The state vector is $\mathbf{x}_t = (\mathbf{x}_t^{(w)\prime}, \mathbf{x}_{t-1}^{0\prime})$ where $\mathbf{x}_t^{(w)}$ is a $m^{(w)} \times 1$ state vector for w_t and $\mathbf{x}_{t-1}^{0\prime} = (x_{t-1}, \ldots, x_{t-d-sD})$ is a $d_0 \times 1$ vector of past observations. Let

$\boldsymbol{\delta}' = (\delta_1, \ldots, \delta_{d_0})$, where $-\delta_j$ are the coefficients of the polynomial $\nabla^d \nabla_s^D$, so that $\nabla^d \nabla_s^D = 1 - \sum_{j=1}^{d_0} \delta_j B^j$. Then the new transition equation is given by (2.4) with

$$\mathbf{T} = \left[\begin{array}{c|cc} \mathbf{T}^{(w)} & \mathbf{0}_{m^{(w)}, d_0} & \\ \mathbf{Z}^{(w)} & \boldsymbol{\delta}' & \\ \mathbf{0}_{d_0-1, m^{(w)}} & \mathbf{I}_{d_0-1} & \mathbf{0}_{d_0-1} \end{array}\right]$$

$$\mathbf{r}' = (\mathbf{r}^{(w)\prime}, \mathbf{0}_{d_0}{}') \tag{2.13}$$

where $\mathbf{Z}_t^{(w)\prime} = (1, \mathbf{0}'_{m^{(w)}-1})$, $\mathbf{T}^{(w)}$ and $\mathbf{r}^{(w)}$ are the same as in (2.10). The new measurement equation, given by (2.5) with $\mathbf{Z}_t' = (\mathbf{Z}_t^{(w)\prime}, \boldsymbol{\delta}')$, is essentially an undifferencing operator.

The Kalman filter is initialized at time $t_0 = d_0$ with $\hat{\mathbf{x}}'_{d_0+1|d_0} = (\mathbf{0}_m{}', \mathbf{x}'^{0}_{d_0})$ and

$$\mathbf{P}_{d_0+1|d_0} = \begin{bmatrix} \mathbf{P}_{1|0}^{(w)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{2.14}$$

where $\mathbf{P}_{1|0}^{(w)}$ is the solution to (2.11) with $\mathbf{T}$ and $\mathbf{R}$ replaced by $\mathbf{T}^{(w)}$ and $\mathbf{R}^{(w)}$ respectively. With no missing data, the likelihood computed from the observations x_t is identical to that computed from the differenced observations w_t. The likelihood is maximized as before, using partial autoregressive and moving average coefficients to ensure stationarity and invertibility of w_t.

Note that this approach requires d_0 consecutive observations at the beginning of the series. If a missing value occurs in the beginning, then the likelihood can be computed by reversing the series (i.e., ordering the data by $x_n, x_{n-1}, \cdots, x_1$) and applying the Kalman filter to the reversed series. If there are missing values at both ends of the series, this approach will not work. In this case, the formulation proposed by Kohn and Ansley (1986)

can be applied: see Section 8 for further discussion.

2.5 Seasonal Models

The extension to the general seasonal ARIMA model given by (2.1) follows from expansion of the autoregressive and moving average ordinary and seasonal operators.

In the stationary case, the state vector $\mathbf{x}_t$ has length $m = \max(p+sP, q+sQ+1)$, and the parameters $\phi_1, \cdots, \phi_p$ and $\theta_1, \cdots, \theta_q$ are replaced by the appropriate coefficients of $\phi(B)\Phi(B)$ and $\theta(B)\Theta(B)$. Stationarity and invertibility are assured by transforming each of the ordinary and seasonal parts.

The nonstationary case is extended to seasonal models in the same fashion.

3. Leave-k-Out Diagnostics for ARIMA Models

In this section, we describe our basic leave-k-out diagnostics for both ARIMA model coefficient estimates, and the innovations variance estimate. As we shall see by example in several simulated series, the diagnostics based on the innovations variance yield much sharper results than those for the coefficients.

3.1 The Basic Leave-k-out Diagnostics

Diagnostics for Coefficients

Denote the maximum likelihood estimate (MLE) of $\boldsymbol{\alpha}$ by $\hat{\boldsymbol{\alpha}}$. Let $A = \{ t_1, t_2, \ldots, t_k \}$ be an arbitrary subset of $\{ 1, 2, \ldots, n \}$, and let $\hat{\boldsymbol{\alpha}}_A$ denote the MLE with observations $y_{t_1}, \cdots, y_{t_k}$ treated as missing. If some of the observations in A have an undue influence on the estimate $\hat{\boldsymbol{\alpha}}_A$, then this will often reveal itself in the form of a substantial difference between $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\alpha}}_A$. We define the ***empirical influence on the coefficients of the subset*** A by

$$\mathrm{EI}(A) = -n(\hat{\boldsymbol{\alpha}}_A - \hat{\boldsymbol{\alpha}}). \tag{3.1}$$

Standardizing by the factor n leads to a non-degenerate asymptotic form for (3.1) (see Appendix B).

In the case of independent observations, one almost always deletes a single observation at a time and computes various diagnostic statistics. However, the time series situation differs from the case of independent observations in at least two important ways: (a) structure is imposed by time ordering, and (b) influential observations often come in the form of an "outlier patch" or other local "structural" change extending over several observations. Leave-one-out diagnostics can fail to give clear evidence of influence in the case of patchy disturbances such as outliers (an example of this is provided below). Such behavior might be regarded as a form of "masking" since the effect of any single outlier in s...

patch can be overwhelmed by the effect of the other outliers. Fortunately, this kind of situation is easily dealt with in time series (unlike as in unstructured independent observation problems) by leaving out k consecutive observations; that is, by taking $A = A_{k,t}$ to consist of the k time points centered at t: $(t-[\frac{k-1}{2}], \cdots, t+[\frac{k}{2}])$, where $[x]$ denotes the largest integer less that or equal to x. To simplify notation, denote $\mathrm{EI}(A_{k,t})$ by $\mathrm{EI}(k,t)$ and $\hat{\alpha}_{A_{k,t}}$ by $\hat{\alpha}_{k,t}$.

For patches at the ends of the series, where $t \le [\frac{k-1}{2}]$ or $t > n-[\frac{k}{2}]$, $\mathrm{EI}(k,t)$ is computed with the patch truncated in the obvious manner. For nonstationary models, the series will be reversed to obtain $\mathrm{EI}(k,t)$ for $t = 1, \cdots, d_0+[\frac{k-1}{2}]$ where d_0 is the order of the differencing (see the comments under nonstationary models in Section 2).

A strategy for determining the largest k that needs to be considered for a given data set will emerge, based on the empirical examples of Section 3.2.

Standardizing **EI**

The empirical influence $\mathrm{EI}(k,t)$ is an r-dimensional vector, and as such is difficult to interpret. Further, the empirical influence is relative, and comparable only within a data set. Hence it is useful, as in the ordinary regression context, to consider a quadratic form diagnostic measure of influence for coefficients

$$DC(k,t) = \mathrm{EI}'(k,t)\, \boldsymbol{M}\, \mathrm{EI}(k,t) \tag{3.2}$$

where $\boldsymbol{M}$ is an appropriate positive semi-definite $r \times r$ matrix. As in the regression setting, it is natural to choose $\boldsymbol{M}$ to be the inverse of covariance matrix of $\hat{\alpha}$.

Although the exact covariance matrix for $\hat{\alpha}$ is not known, it can be approximated by the asymptotic information matrix $\mathrm{I}(\alpha)$. It is well known that $\hat{\alpha}$ is asymptotically normal under regularity conditions (see, for example, Fuller, 1976):

$$\sqrt{n}\,(\hat{\alpha}-\alpha) \;\rightarrow\; N_r(0, \mathrm{I}(\alpha)^{-1})\,. \tag{3.3}$$

If $\hat{\mathbf{I}}(\boldsymbol{\alpha})$ is a consistent estimator of $\mathbf{I}(\boldsymbol{\alpha})$, then the Mann-Wald theorem implies that

$$n(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})'\,\hat{\mathbf{I}}(\boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}) \;\rightarrow\; \chi_r^2 \tag{3.4}$$

where χ_r^2 denotes a chi-square random variable with r degrees of freedom. Thus, it is natural to choose $\mathbf{M}$ to be $n^{-1}\hat{\mathbf{I}}(\boldsymbol{\alpha})$.

One estimator of $\mathbf{I}(\boldsymbol{\alpha})$ is $\mathbf{I}(\hat{\boldsymbol{\alpha}})$, the expected information evaluated at the maximum likelihood estimate. Although not commonly available in the literature, a closed form expression for $\mathbf{I}(\boldsymbol{\alpha})$ in terms of $\boldsymbol{\alpha}$ exists (this expression is derived in Appendix A). Using this expression, we take as our leave-k-out diagnostic for coefficients

$$\begin{aligned} DC(k,t) &= \frac{1}{n}\,\mathbf{EI}'(k,t)\,\mathbf{I}(\hat{\boldsymbol{\alpha}})\mathbf{EI}(k,t) \qquad (3.5) \\ &= n(\hat{\boldsymbol{\alpha}}-\hat{\boldsymbol{\alpha}}_{k,t})'\,\mathbf{I}(\hat{\boldsymbol{\alpha}})(\hat{\boldsymbol{\alpha}}-\hat{\boldsymbol{\alpha}}_{k,t})\,. \end{aligned}$$

Although the distribution of $DC(k,t)$ is not known, the use of the χ_r^2 distribution allows one to view $DC(k,t)$ on a familiar scale. This corresponds to using the F distribution as a reference for Cook's Distance (Cook and Weisberg, 1982) and DFFITS (Belsley et. al., 1980). The χ_r^2 distribution is used in the time series case, rather than an F distribution, since $\mathbf{I}(\boldsymbol{\alpha})$ does not involve the nuisance parameter σ^2.

Following previous applications of leave-k-out diagnostics, we recommend judging a point or patch of points to be influential if the "p-value" of $DC(k,t)$ based on the χ_r^2 reference distribution is smaller than .5. Empirical evidence shows this guideline is quite useful, except near the region of noninvertibility or nonstationarity.

Diagnostics for the Innovations Variance

The influence of a subset A can also be measured by evaluating the effect of its removal on the MLE of the innovations variance estimate $\hat{\sigma}^2$. The innovations variance is a nuisance parameter, and at first thought it might therefore appear to have less intuitive appeal as a basis for leave-k-out diagnostics than the coefficient estimates $\boldsymbol{\alpha}$. However, it turns out that a diagnostic based on the innovations variance leads to a more effective tool than DC for identifying outliers. The diagnostic is formed in the same manner as above: A standardized version of $-(\hat{\sigma}^2_{k,t} - \hat{\sigma}^2)$ is computed, where $\hat{\sigma}^2_{k,t}$ is the MLE of σ^2 with observations at times $t \in A_{t,k}$ treated as missing.

Again, the standardization is based on asymptotic theory. Under regularity conditions, $\hat{\sigma}^2$ is asymptotically independent of $\hat{\alpha}$, and

$$\sqrt{n}\,(\hat{\sigma}^2 - \sigma^2) \rightarrow N(0, 2\sigma^4)\ . \tag{3.6}$$

If $\tilde{\sigma}^2$ is a consistent estimate of σ^2, then by the Mann-Wald theorem

$$\frac{n}{2}\left[\frac{\hat{\sigma}^2}{\tilde{\sigma}^2} - 1\right]^2 \rightarrow \chi_1^2\ .$$

Thus, we propose to use as leave-k-out diagnostic for innovations variance

$$DV(k,t) = \frac{n}{2}\left[\frac{\hat{\sigma}^2}{\hat{\sigma}^2_{k,t}} - 1\right]^2 \tag{3.7}$$

with the reference distribution being a chi-squared with one degree of freedom (χ_1^2). Again, one suspects an observation y_t to be influential if the p-value for $DV(k,t)$ is less than .5 using a χ_1^2.

Relationship Between DV and Fox Tests for AO

The difference $n[\hat{\sigma}^2 - \hat{\sigma}_{k,T}^2]$ is asymptotically equivalent to the squared interpolated residual $(x_t - \hat{x}_t^n)^2$ where $\hat{x}_t^n = E(x_t \mid x_l, t = 1, \ldots, n, l \neq t)$. The *interpolated* residual was used by Fox (1972) as a basis for testing the presence of a (parametric) AO type outlier at a fixed time t. At first glance this equivalence may seem surprising since $(\hat{\sigma}^2 - \hat{\sigma}_{k,T}^2)$ is based on the *prediction* residuals $e_t = x_t - \hat{x}_t$ and not on the interpolation residuals $x_t - \hat{x}_t^n$. But, by using the smoothing form of the likelihood (Schweppe, 1973), it is easy to show the claimed asymptotic equivalence.

It is important to note that in spite of the asymptotic equivalence, the finite sample differences can be significant, e.g., $\hat{\sigma}^2 - \hat{\sigma}_{k,T}^2$ can be negative, whereas $(x_t - \hat{x}_t^n)^2$ cannot.

3.2 Outlier Models and Examples

Outlier Models

In the following examples, we focus on influential points caused by *outliers*. Influential observations may also be the result of structural changes, such as level shifts or variance changes. We shall discuss application of leave-k-out diagnostics to such problems in Sections 7 and 8.

We examine the performance of DC and DV under two types of contamination commonly used in other studies (see, for example, Fox, 1972; Denby and Martin, 1979; Martin and Yohai, 1986; Tsay, 1986): the additive outliers (AO) model and the innovations outliers (IO) model. The primary focus will be on AO models.

Let x_t be a Gaussian ARIMA process specified by (2.1). Then y_t behaves according to a *constant magnitude* AO model if

$$y_t = x_t + \zeta z_t \tag{3.8}$$

where ζ is constant and z_t is a fixed 0–1 process. The magnitude of the outliers is ζ; isolated outliers and patches are created by appropriate choice of 0's and 1's for z_t.

A *constant magnitude* IO model is formed through contamination in the innovations process ε_t. Let ε_t be a contaminated white noise process, with

$$\varepsilon_t = \gamma_t + \zeta z_t \tag{3.9}$$

where γ_t are independent Gaussian random variables with zero mean and variance σ^2, and ζ, z_t are as above. Then y_t follows an ARIMA IO model if it is generated by (2.1) with the ε_t given by (3.9).

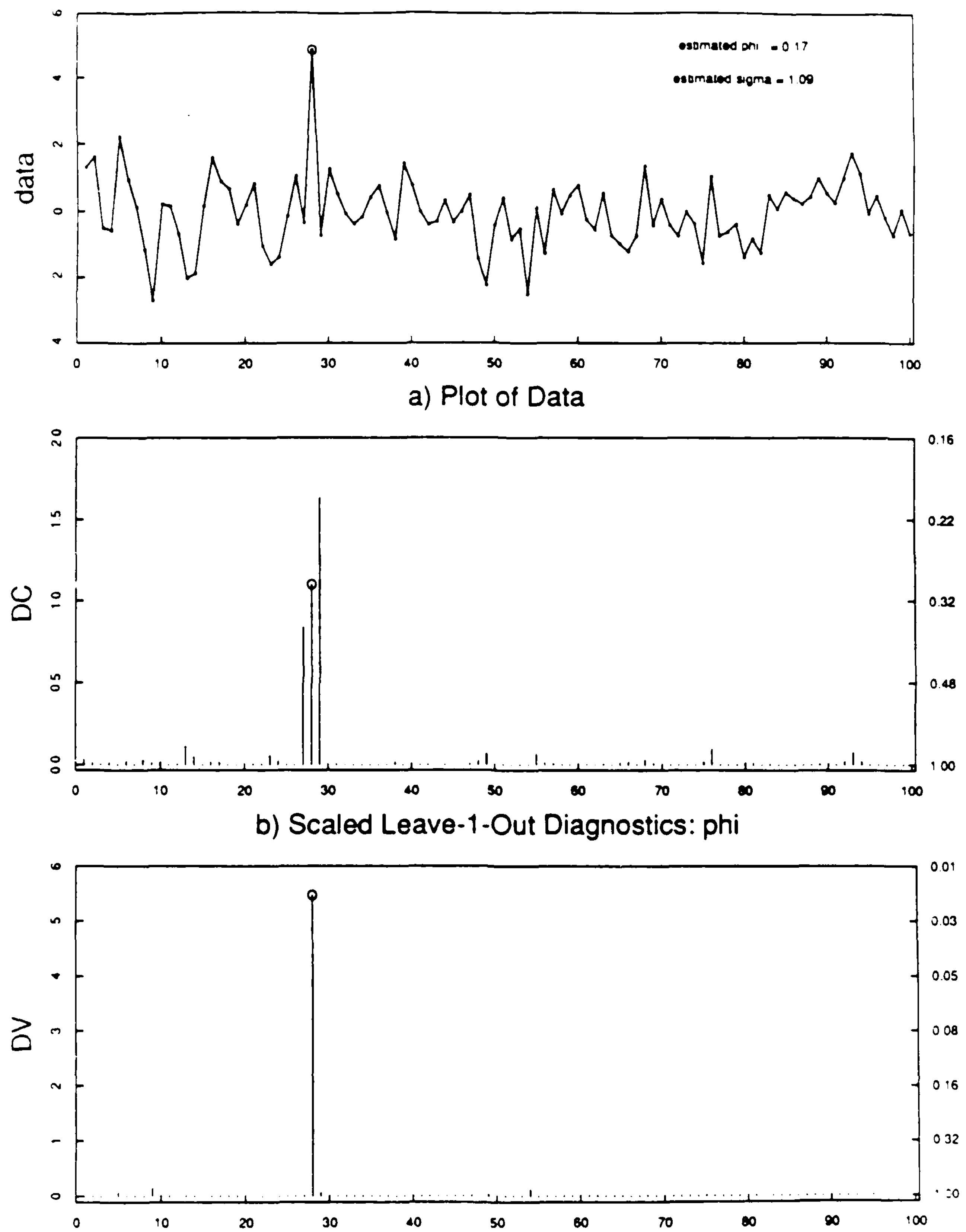
The models (3.8) and (3.9) are actually somewhat special AO and IO forms. More general AO and IO (and other) outlier models for time series are possible (see Martin and Yohai, 1986, for a very flexible "general replacement" model).

Example 3.1: AR(1), $\phi = .4$, $\sigma^2 = 1$, AO model with 1 isolated outlier

Starting with a simple case, we examine a simulated AR(1) series of 100 points from Gaussian white noise with $\phi = .4$ and a single additive outlier of +4 at point 28. The MLE fit of an AR(1) model with the entire series yielded $\hat{\phi} = 0.17$ and $\hat{\sigma}^2 = 1.09$. The data is plotted in Figure 1a and the outlier is marked by "o". The leave-1-out diagnostics, $DC(1,\cdot)$ and $DV(1,\cdot)$, for $\hat{\phi}$ and $\hat{\sigma}^2$ are displayed in Figures 1b and 1c. The p-values corresponding to a χ_1^2 distribution are displayed on the right axis. The p-values for $DC(1,t)$ at $t = 27, 28, 29$ are all smaller than .5, while the p-value for $DC(1,t)$ is considerably greater tha .5 for all other times. Thus y_{27}, y_{28}, and y_{29} are judged to be influential. By contrast, only $DV(1,28)$ is significant and has the much smaller p-value of about .02. This example is indicative of a general pattern which we establish analytically in Section 4: an outlier is smeared across several values of $DC(1,\cdot)$, but is identified exactly by $DV(1,\cdot)$. In particular, the smearing for $DC(1,\cdot)$ extends by one time unit in each direction from $t = 28$.

Example 3.2: AR(1), $\phi = .4$, $\sigma^2 = 1$, IO model with 1 isolated outlier

This is the same series as in Example 3.1, except that the outlier at point 28 is of the IO type. The MLE's of the parameters are: $\hat{\phi} = .27$ and $\hat{\sigma}^2 = 1.06$. Figures 2a, 2b, and 2c display the data and leave-1-out diagnostics $DC(1,\cdot)$ and $DV(1,\cdot)$. The problem of smearing for DC is considerably worse than in Example 1. The diagnostic is significant only at time $t = 27$, while the p-value of about .7 at $t = 28$ is quite insignificant. However, there is no smearing with $DV(1,\cdot)$: $DV(1,28)$ is several magnitudes larger than $DV(1,t)$ for any other t, and hence DV identifies the outlier. Again, this is a general behavior, established in Section 4: DC is large at time points just prior to the occurrence of an isolated IO type outlier, but is small at the time of occurrence of the outlier, while DV is large only at the time of occurrence of the outlier.



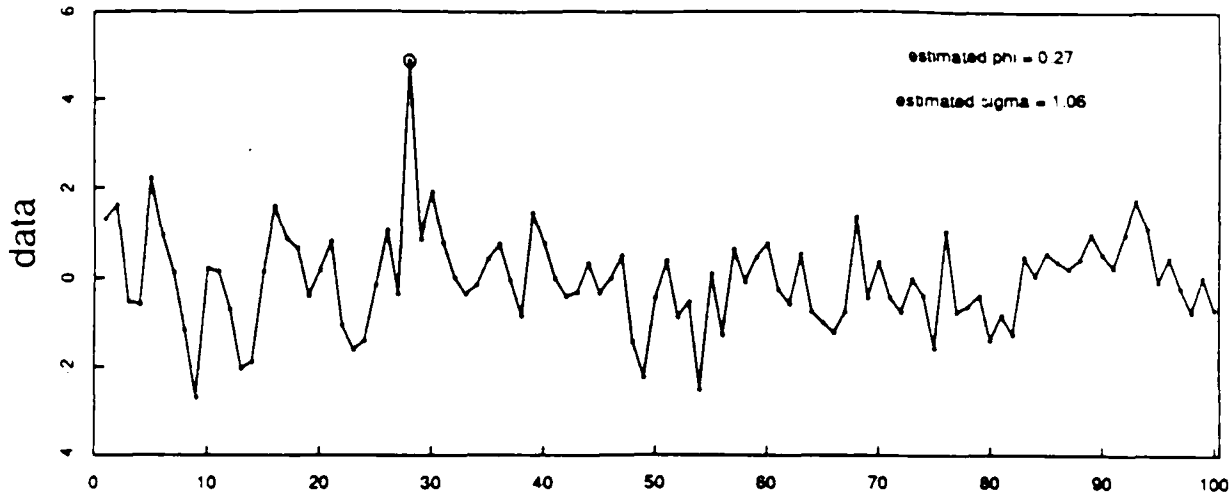
a) Plot of Data

b) Scaled Leave-1-Out Diagnostics: phi

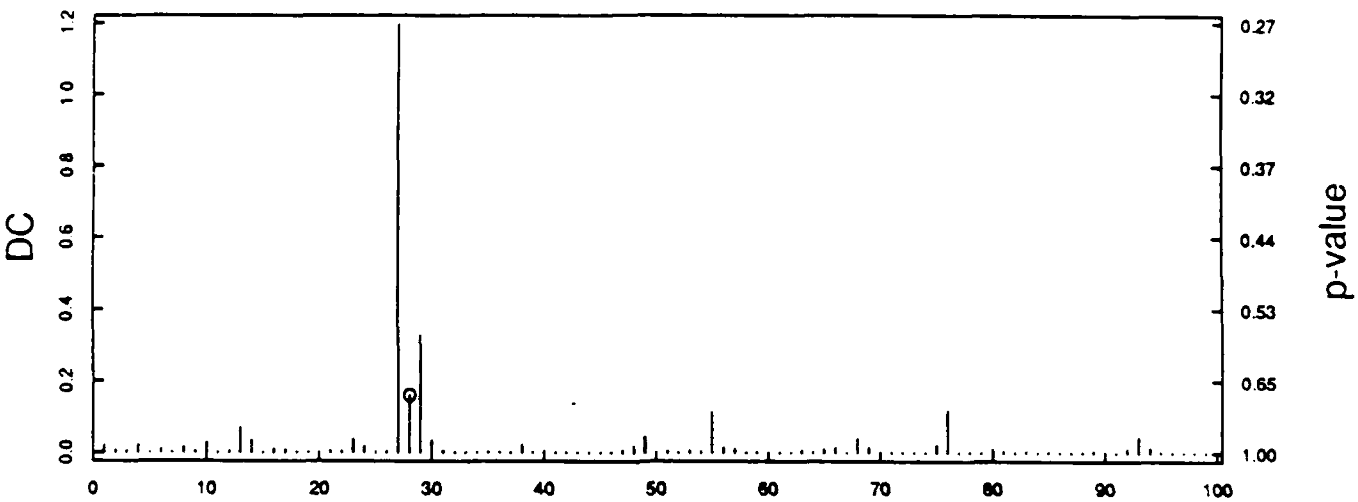
c) Scaled Leave-1-Out Diagnostics: Innovations Variance

Example 3.1: Simulated AR(.4) With 1 Isolated Additive Outlier

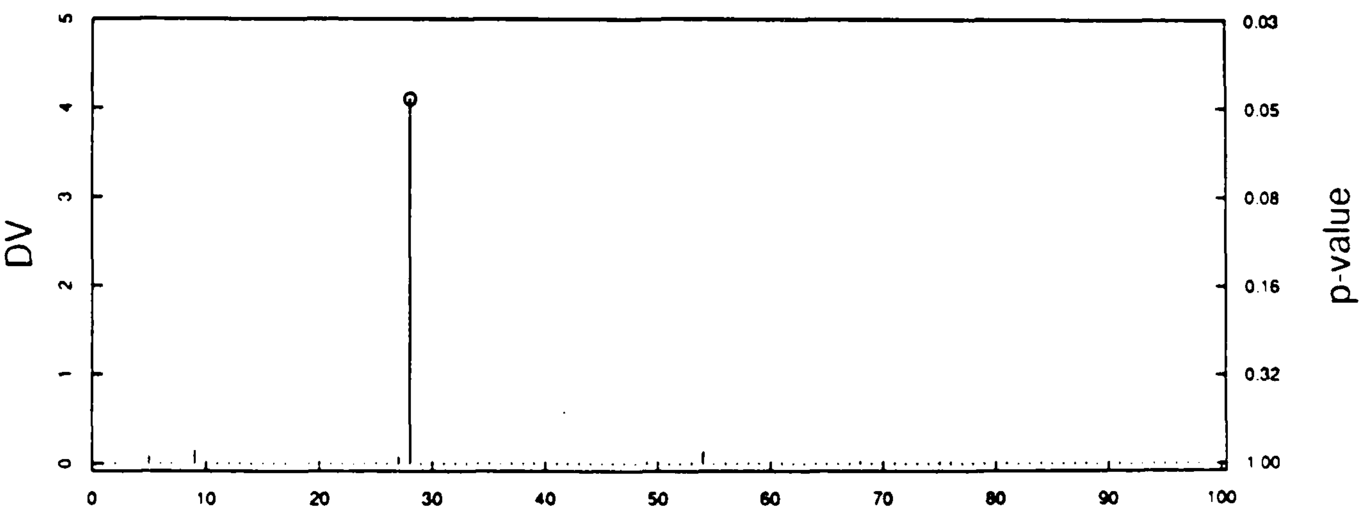
Figure 1



a) Plot of Data



b) Scaled Leave-1-Out Diagnostics: phi



c) Scaled Leave-1-Out Diagnostics: Innovations Variance

Example 3.2: Simulated AR(.4) With 1 Isolated Innovations Outlier

Figure 2

In all examples to follow, plots of DC are omitted for simplicity. However, in Section 5, we discuss the use of DC as a tool to determine whether an outlier is AO or IO.

Example 3.3: MA(1), $\theta = -.5$, $\sigma^2 = 1$, *AO model with 1 patch and 1 isolated outlier*

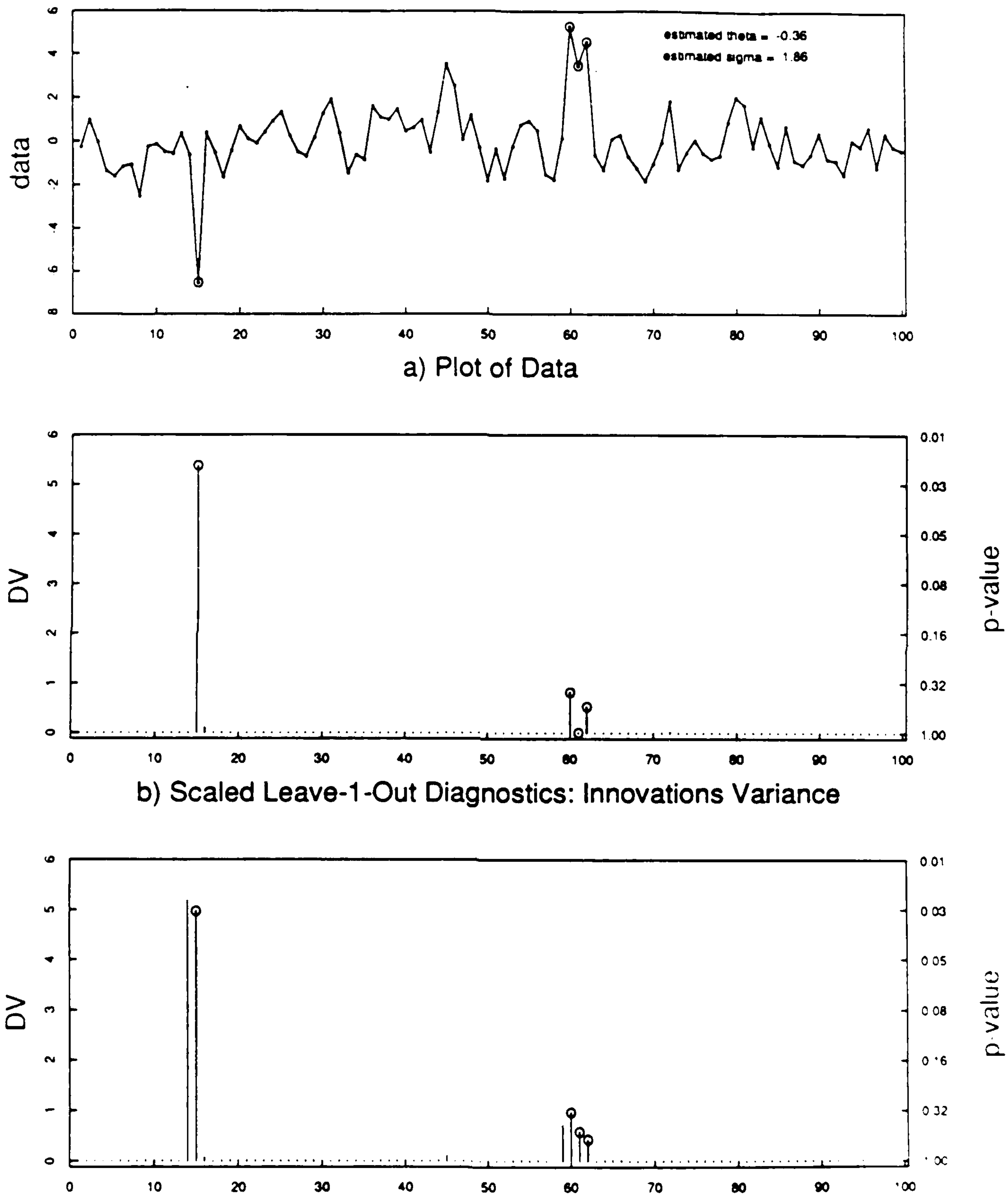
This example is a simulated MA(1) series with $\theta = -.5$ and both a patch of three outliers of size +4 at points 60–62 and an isolated outlier of size -4 at time 15. The outliers are all of the AO type. Figure 3a shows the data; the MLE's are $\hat{\theta} = -.36$ and $\hat{\sigma}^2 = 1.86$. Leave-1-out through leave-4-out diagnostics for DV are displayed in Figures 3b–3e.

Recall that for $k \geq 2$, $DV(k,t)$ represents the influence of a patch of k observations centered at t. For even k, t is the closest point to the left of the "center" of the patch. For example, with $k = 2$, $DV(k,t)$ corresponds to the diagnostic computed when y_t and y_{t+1} are left out.

Although, leave-1-out diagnostics clearly identify the isolated outlier, there is just barely an indication (use the p-value of $.5$ as a guideline) of something going on at $t = 60$ and 62. Leave-1-out is not adequate for detecting the patch of outliers. Leaving a single point out in the patch is insufficient because the remaining outliers in the patch comprise the bulk of the influence of the patch. Leave-2-out and leave-3-out provide progressively stronger evidence of the patch of outliers. The value $DV(3,61)$, is over five times larger than other neighboring diagnostic values.

3.3 Patch Length Determination Strategy

Note that the isolated outlier in Figure 3 is smeared in the leave-k-out diagnostics for $k = 2, 3, 4$. For k=2, both $DV(2, 14)$ and $DV(2, 15)$ are highly significant, and have nearly the same value as $DV(1, 15)$. Similar behavior is observed for $k = 3$ and $k = 4$. The general pattern is as follows: $k-1$ values of $DV(k, \cdot)$ surrounding the location of an isolated outlier at t_0 are significant, and have nearly the save value as $DV(k, t_0)$! This corresponds



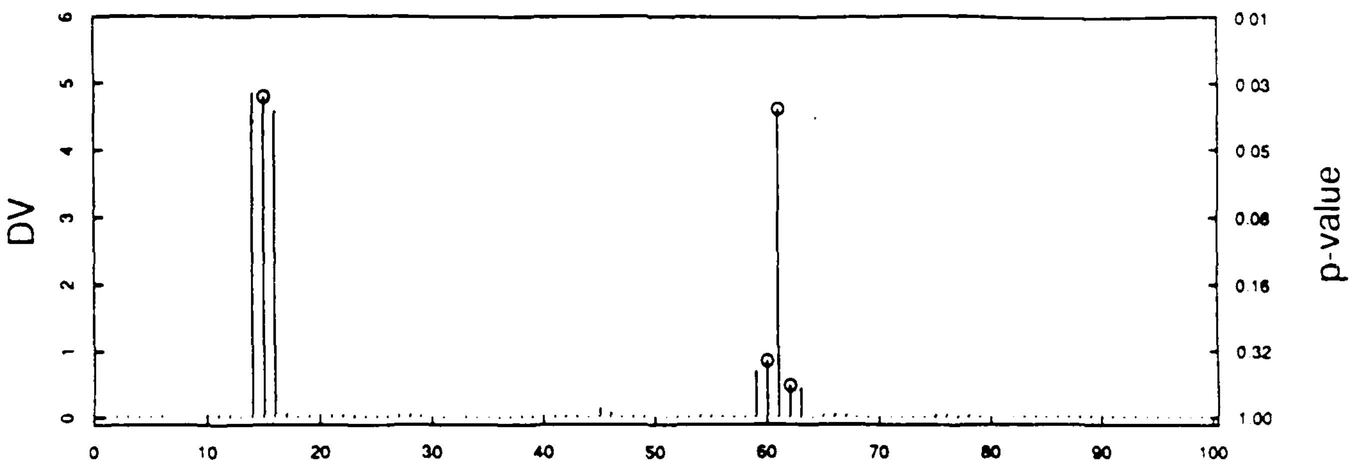
a) Plot of Data

b) Scaled Leave-1-Out Diagnostics: Innovations Variance

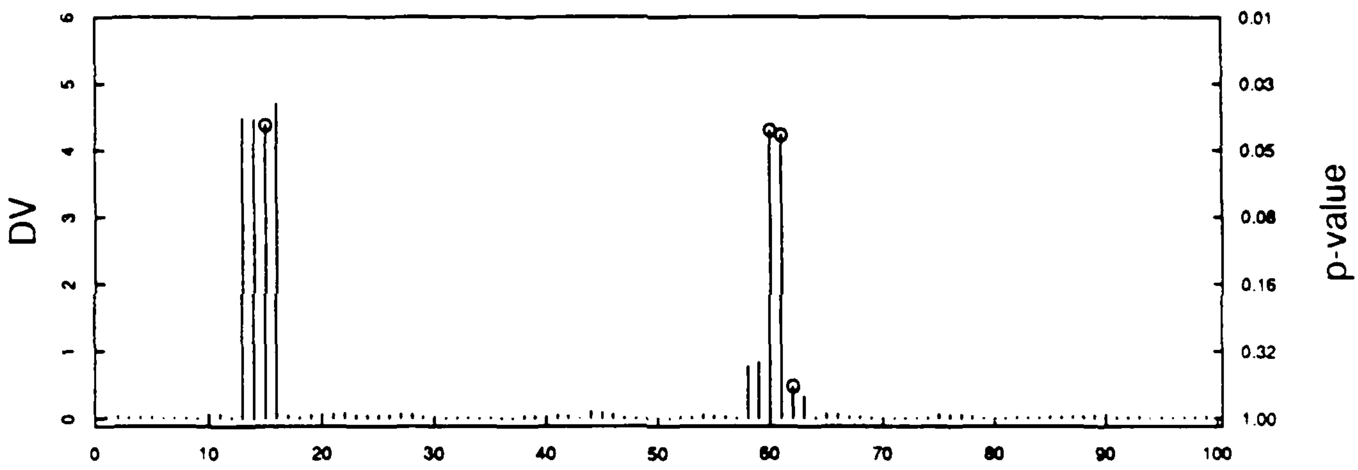
c) Scaled Leave-2-Out Diagnostics: Innovations Variance

Example 3.3: Simulated MA(-.5) With 1 Patch and 1 Isolated Additive Ou

Figure 3



d) Scaled Leave-3-Out Diagnostics: Innovations Variance



e) Scaled Leave-4-Out Diagnostics: Innovations Variance

Example 3.3: Simulated MA(-.5) With 1 Patch and 1 Isolated Additive Outlier

Figure 3

to what one might intuitively expect for an isolated outlier: deletion of a patch which includes an isolated outlier has nearly the same affect as deleting only the isolated outlier

Similar behavior occurs for a patch of outliers. For example, $DV(4, t)$ yields values at $t = 60$ and $t = 61$ which are nearly equal to $DV(3, 61)$. In general, for a patch of k_0 outliers centered at t_0, the following property holds:

- For $k \geq k_0$, there are $k - k_0 + 1$ subsets $A_{k,t}$ which completely overlap the patch, and for deletion of these subsets, the magnitude of $DV(k, t)$ is roughly the same and significant (i.e., the associated p-value is less than .5).

Thus, we judge an influential patch to be of length $k_0 \geq 1$ centered at t_0 if $DV(k_0, t_0)$ is significant, and the above property holds. If $DV(k_0, t_0)$ is significant and the above property fails to hold, then this is an indication that a broader patch of outliers is present.

This provides us with an initial strategy for identifying patches of influential points: Compute leave-k-out diagnostics for increasing $k = 1, 2, \ldots,$ until the magnitude of $DV(k, t)$ does not "significantly" increase for any t. The length of a patch will be estimated as one less than the first value of k for which nearly uniform smearing is in evidence. We shall improve this strategy by incorporating iterative deletion in Section 6.

We close this section with a caveat: Diagnostics in MA models often exhibit different characteristics than in AR models. In particular, the smearing for DC is slightly worse than the AR(1) case. Also, MA models are susceptible to ''start-up'' effects: outliers at the ends of a series are subject to more smearing. These features correspond to the fact that MA models have infinite autoregressive representations.

4. Smearing and the Expected Diagnostic

The examples of Section 3 have revealed a major difference between leave-k-out coefficients diagnostics for time series (including $k = 1$) and the usual regression coefficients diagnostics for independent data. Namely, there is a *smearing* of the effect of an isolated outlier to adjacent points. A given point may be judged influential because of an outlier at an adjacent point. Hence, interpretation of leave-k-out diagnostics for coefficients is not so clear as in the usual regression case. On the other hand, diagnostics for the innovations variance in Section 3 displayed much smaller, and often negligible, smearing effects. In this section, we use an asymptotic approximation to establish an analytical rationale for these different smearing effects. Although the approach will work for any ARIMA model, the computations are quite tedious for all but low order models (where they are also tedious). Thus, after introducing the general expressions for the limiting forms of the diagnostics in Sections 4.1 and 4.2, we concentrate on obtaining explicit calculations for the AR(1) case in Sections 4.3 and 4.4.

4.1 Expected Asymptotic Diagnostic for Coefficients

In order to understand the smearing behavior of the diagnostic (3.2) for coefficients, it is helpful to use an asymptotic representation of $DC(A)$ for general subset deletions A. For subsets $A_{k,t}$ of fixed size k one usually has $\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_{k,t} = 0(\frac{1}{n})$, and correspondingly we are interested in the asymptotic behavior of $n \cdot DC(A)$. Since the asymptotic distribution of this quantity is quite complicated, we work with the *expected asymptotic diagnostic* for the coefficients

$$
\begin{aligned}
EDC(A) &= E\,[\lim_{n \to \infty} n\, DC(A)] \\
&= E\,[\lim_{n \to \infty} n^2 (\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_A)' \mathbf{I}(\hat{\boldsymbol{\alpha}}) (\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_A)] .
\end{aligned} \tag{4.1}
$$

Explicit computations of (4.1) are based on one-step approximations to $\hat{\alpha}_A$ by $\hat{\alpha}$. Denote the (efficient) *score function* $\Psi_n(\alpha)$ and denote its derivative $\dot{\Psi}_n(\alpha)$:

$$\Psi_n(\alpha) = \frac{\partial \log L(\alpha; \mathbf{y}_n)}{\partial \alpha} \qquad \dot{\Psi}_n(\alpha) = \frac{\partial^2 \log L(\alpha; \mathbf{y}_n)}{\partial \alpha \, \partial \alpha'}$$

where the log-likelihood $L(\alpha; \mathbf{y}_n)$ is given by (2.8) with $\mathbf{x}_n$ replaced by $\mathbf{y}_n$ and σ^2 suppressed. Let $\log L^{(A)}(\alpha; \mathbf{y}_n)$ be the log-likelihood with subset A removed, and let $\Psi_n^{(A)}(\alpha)$ and $\dot{\Psi}_n^{(A)}(\alpha)$ denote the corresponding score function and its derivative. Under suitable regularity conditions, a Taylor series expansion of $\Psi_n^{(A)}(\hat{\alpha}_{(A)})$ about $\hat{\alpha}$ yields

$$0 = \Psi_n^{(A)}(\hat{\alpha}) + (\hat{\alpha}_A - \hat{\alpha})' \left(\dot{\Psi}_n^{(A)}(\hat{\alpha}) + o_p(n) \right) . \tag{4.2}$$

One difference between (4.2) and the usual log-likelihood expansion is that scaling $\hat{\alpha}_{(A)} - \hat{\alpha}$ by n (rather than $\sqrt{n}$) leads to a non-degenerate asymptotic form.

Since $\hat{\alpha} \rightarrow_p \alpha$, $-\frac{1}{n} \dot{\Psi}_n^{(A)}(\alpha) \rightarrow_p I(\alpha)$ and $\Psi_n(\hat{\alpha}) = 0$, we may rearrange (4.2) to obtain

$$\begin{aligned} EI(A) &= -n(\hat{\alpha}_A - \hat{\alpha}) \\ &= (-I(\hat{\alpha})^{-1} + o_p(1)) \Psi_n^{(A)}(\hat{\alpha}) \\ &= -I(\hat{\alpha})^{-1} (\Psi_n^{(A)}(\hat{\alpha}) - \Psi_n(\hat{\alpha})) + o_p(1) \\ &= -I(\alpha)^{-1} (\Psi_n^{(A)}(\alpha) - \Psi_n(\alpha)) + o_p(1) . \end{aligned} \tag{4.3}$$

Combining (4.3) and (3.5) gives the asymptotic form of $n\,DC(A)$ for general subset deletions A :

$$n\,DC(A) = \Delta(A, \alpha)' I(\alpha)^{-1} \Delta(A, \alpha) + o_p(1) \tag{4.4}$$

where

$$\Delta(A, \alpha) \equiv \Psi_n^{(A)}(\alpha) - \Psi_n(\alpha) . \tag{4.5}$$

Hence, the expected asymptotic diagnostic for coefficients is given by

$$EDC(A) = E\,[\lim_{n\to\infty} \Delta(A,\alpha)\, I\,(\alpha)^{-1}\Delta(A,\alpha)\,] \,. \tag{4.6}$$

Our problem is now reduced to computing the difference $\Delta(A,\alpha)$ between the score function with and without subset A included, and evaluating the expectation in (4.6).

4.2 Expected Asymptotic Diagnostic for Innovations Variance

In the same spirit as in (4.1), we shall use the *expected asymptotic diagnostic* for the innovations variance

$$EDV(A) = E\,[\lim_{n\to\infty} n\; DV(A)]\,.$$

Denote the score function for σ^2 and its derivative by $\Psi_n(\sigma^2)$ and $\dot\Psi_n(\sigma^2)$, respectively, and denote these functions with subset A removed by $\Psi_n^{(A)}(\sigma^2)$ and $\dot\Psi_n^{(A)}(\hat\sigma^2)$. Then

$$n\,(\hat\sigma_A^2 - \hat\sigma^2) = \left[\frac{1}{n}\dot\Psi_n^{(A)}(\hat\sigma^2)\right]^{-1} \Psi_n^{(A)}(\hat\sigma^2) + o_p(1) \tag{4.7}$$

where $\left[\frac{1}{n}\dot\Psi_n^{(A)}(\hat\sigma^2)\right]^{-1} = 2\hat\sigma_A^4 + o_p(1) = 2\sigma^4 + o_p(1)$. From this and the definition (3.7), applied to general subset deletions A, we have

$$\begin{aligned} n\,DV(A) &= \left[\frac{1}{2\hat\sigma_A^4}\right] n^2 (\hat\sigma^2 - \hat\sigma_A^2)^2 \\ &= 2\,\hat\sigma_A^4\,[\Psi_n^{(A)}(\hat\sigma^2)]^2 + o_p(1) \\ &= (2\sigma^4)(\Psi_n^{(A)}(\sigma^2) - \Psi_n(\sigma^2))^2 + o_p(1)\,. \end{aligned}$$

Then with

$$\Delta(A;\sigma^2) = \Psi_n^{(A)}(\sigma^2) - \Psi_n(\sigma^2) \tag{4.8}$$

we have

$$EDV(A) = 2\sigma^4 E\,(\lim_{n\to\infty} \Delta^2(A;\sigma^2)), \tag{4.9}$$

and our problem is reduced to computation of $\Delta(A;\sigma^2)$, and evaluating the expectation in (4.9).

4.3 AO Models: AR(1) Case

Expressions for $EDC^{AO}(t)$

We now compute $EDC(t) = EDC(A)|_{A=t}$ and $EDV(t) = EDV(A)|_{A=t}$ for the AR(1) model with AO type outliers. Let x_t denote an outlier free Gaussian process. Straightforward algebra (see Appendix B) shows that *for the outlier free process*, the difference in the score functions for ϕ with and without x_t is given by

$$\Delta(t;\phi) = -\frac{\phi}{1+\phi^2} - \frac{1}{\sigma^2}\Big[-\phi x_t^2 + x_t(x_{t-1}+x_{t+1}) - \frac{\phi}{(1+\phi^2)^2}(x_{t-1}+x_{t+1})^2\Big]. \tag{4.10}$$

A pleasant feature of (4.10) is that $\Delta(t;\phi)$, and hence $EDC(t)$, depends only on x_{t-1}, x_t, x_{t+1}. More generally, for AR(p) models, $\Delta(t,\alpha) = \Psi_n^{(t)}(\alpha) - \Psi_n(\alpha)$, depends only on $x_{t-p}, x_{t-p+1}, \ldots, x_{t+p}$. Replacing x_t by y_t in (4.10), we can derive the difference in the score functions for various outlier models.

First consider the case where y_t is observed with a single AO type outlier at t_0; i.e., y_t behaves according to (3.8) with $z_t = 1$ only at $t = t_0$:

$$y_t = \begin{cases} x_t & t \neq t_0 \\ x_t + \zeta & t = t_0 \ . \end{cases} \tag{4.11}$$

The difference in the score function $\Delta(t_0;\phi)$ when we leave y_{t_0} out, for example, is given by (4.10) with x_{t_0} replaced by $x_{t_0}+\zeta$. Similarly, for leaving y_{t_0+1} out we evaluate $\Delta(t_0+1,\phi)$ by setting $t=t_0+1$ in (4.10) and again replacing x_{t_0} with $x_{t_0}+\zeta$. The expected asymptotic diagnostics are then computed by substituting the appropriate expressions in (4.6) and taking expectations.

We begin by computing $EDC(t)$ at the location $t=t_0$ of the additive outlier. Using the notation $\Delta^{AO}_{(\zeta;t_0)}(t;\phi)$ in place of $\Delta(t;\phi)$ when $\Delta(t;\phi)$ is computed under an AO model with an outlier of size ζ at time t_0, we have

$$\Delta^{AO}_{(\zeta;t_0)}(t_0;\phi) = -\frac{\phi}{(1+\phi^2)} - \frac{1}{\sigma^2}\Big[-\phi(x_{t_0}+\zeta)^2 + (x_{t_0}+\zeta)(x_{t_0-1}+x_{t_0+1}) - \frac{\phi}{(1+\phi^2)^2}(x_{t_0-1}+x_{t_0+1})^2\Big] \tag{4.12}$$

For the AR(1) model, $I(\phi) = (1-\phi^2)^{-1}$. Substituting this and (4.12) into (4.6), along with some tedious algebra (see Appendix B), yields

$$EDC^{AO}_{(\zeta;t_0)}(t_0) = \left[\frac{\zeta}{\sigma}\right]^4 \phi^2(1-\phi^2) + \left[\frac{\zeta}{\sigma}\right]^2 2(1-\phi^2) + \frac{2(1-\phi^2)}{(1+\phi^2)^2} \tag{4.13}$$

where the notation $EDC^{AO}_{(\zeta;t_0)}(t)$ parallels that for $\Delta^{AO}_{(\zeta;t_0)}(t;\phi)$.

To examine the effects of smearing, we now compute $EDC^{AO}_{(\zeta;t_0)}(t)$ for $t\neq t_0$. It is shown in Appendix B that, as one might expect, $EDC^{AO}_{(\zeta;t_0)}(t)$ is given by the right-hand side of (4.13) with $\zeta=0$ for $t\neq t_0-1, t_0, t_0+1$, which is the expected diagnostic for the noise free process. Thus it suffices to compute the expected diagnostics for $t=t_0-1$ and $t=t_0+1$. Furthermore, by inspection of (4.10), it is evident that the effect of an outlier at t_0 is symmetric: $EDC^{AO}_{(\zeta;t_0)}(t_0-1) = EDC^{AO}_{(\zeta;t_0)}(t_0+1)$. Hence, we need only concern

ourselves with $EDC^{AO}_{(\zeta;t_0)}(t_0+1)$, and computations similar to those above yield

$$EDC^{AO}_{(\zeta;t_0)}(t_0+1) = \left[\frac{\zeta}{\sigma}\right]^4 \frac{\phi^2(1-\phi^2)}{(1+\phi^2)^4} + \left[\frac{\zeta}{\sigma}\right]^2 \left[1-8\frac{\phi^4}{(1+\phi^2)^3}\right] + \frac{2(1-\phi^2)}{(1+\phi^2)^2} . \tag{4.14}$$

Expressions for $EDV^{AO}(t)$

To obtain $EDV^{AO}_{(\zeta;t_0)}(t)$ under the AO model we compute the difference in the score functions $\Delta(t;\sigma^2)$ for σ^2 given by (4.8) with $A = t$ and apply (4.9). Using the same notation as in (4.12) we have for $t = t_0$ (see Appendix C)

$$\Delta^{AO}_{(\zeta;t_0)}(t_0;\sigma^2) = \frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}\Big[-(1+\phi^2)(x_{t_0}+\zeta)^2 + 2\phi(x_{t_0}+\zeta)(x_{t_0-1}+x_{t_0+1}) - \frac{\phi^2}{1+\phi^2}(x_{t_0-1}+x_{t_0+1})^2\Big] \tag{4.15}$$

and (again with tedious algebra)

$$EDV^{AO}_{(\zeta;t_0)}(t_0) = \left[\frac{\zeta}{\sigma}\right]^4 \frac{(1+\phi^2)^2}{2} + \left[\frac{\zeta}{\sigma}\right]^2 8(1+\phi^2) + 1 . \tag{4.16}$$

As with $EDC^{AO}_{(\zeta;t_0)}(t)$, for $t \neq t_0-1,\ t_0,\ t_0+1$, $EDV^{AO}_{(\zeta;t_0)}(t)$ is simply the right-hand side of (4.16) with $\zeta=0$. Also, the same symmetry relations hold for $EDV^{AO}_{(\zeta;t_0)}(t)$, so that $EDV^{AO}_{(\zeta;t_0)}(t_0-1) = EDV^{AO}_{(\zeta;t_0)}(t_0+1)$, where straightforward computations give

$$EDV^{AO}_{(\zeta;t_0)}(t_0+1) = \left[\frac{\zeta}{\sigma}\right]^4 \frac{\phi^4}{2(1+\phi^2)^2} + \left[\frac{\zeta}{\sigma}\right]^2 \frac{2\phi^2}{(1+\phi^2)} + 1 . \qquad (4.17)$$

Comparisons of $EDC^{AO}(t)$ and $EDV^{AO}(t)$

Although the dominant terms in (4.13), (4.14), (4.16), and (4.17) are proportional $\left[\frac{\zeta}{\sigma}\right]^4$, the coefficients for this term are uniformly smaller for (4.17) than for (4.14) relative to (4.16) and (4.13). However, the difference between the behavior of $EDC^{AO}(t)$ and $EDV^{AO}(t)$ is best seen graphically.

Figure 4a plots $EDC^{AO}_{(+4;t_0)}(t)/100$ curve (i.e., the expected asymptotic diagnostic assuming an outlier of size 4 at time t_0) for $t = t_0-3,\ t_0-2,\ \cdots,\ t_0+3$ with $\phi = .3,\ .6,\ .9$. Figure 4b gives the corresponding plot for $EDV^{AO}_{(+4;t_0)}(t)/100$. The scaling factor of $1/100$ approximates the expected value of the diagnostics for a sample size of 100. The asymptotic approximations verify what was observed in Example 1 for AO models: the smearing is worse for DC, and DV tends to be more sensitive.

Due to sampling fluctuation, the patterns of diagnostics observed in Example 1 differ from the expected diagnostics in two regards: the magnitude of DC and DV is larger than EDC and EDV, and the pattern over time for DC is not the same in that the largest diagnostic is for the time point after the outlier ($t = 28$).

In Figure 5a, we compare the amount of smearing graphically for DC and DV as a function of ϕ. The ratios $\dfrac{EDC^{AO}_{(+4;t_0)}(t_0-1)}{EDC^{AO}_{(+4;t_0)}(t_0)}$ (solid line) and $\dfrac{EDV^{AO}_{(+4;t_0+1)}(t_0-1)}{EDV^{AO}_{(+4;t_0)}(t_0)}$ (dashed line) represent the proportional amount of smearing for an outlier of size 4 at t_0 These ratios are always less than unity. However, the expected asymptotic smearing for DV is small in absolute terms for all ϕ, and also substantially smaller than that of DC for all but quite large values of ϕ. The smearing for DC is greater than .5 for a large range of ϕ

values, and this suggests that in such situations, smearing may lead to some confusion when examining DC.

The potential for confusion in fact becomes unquestionably serious in situations where there is more than one outlier present. We demonstrate this in the very simplest context. Suppose y_t is observed with two isolated AO type outliers of size ζ at times $t_0 - 1$ and $t_0 + 1$. Note that for $t \neq t_0$, the expected asymptotic diagnostics $EDC^{AO}_{(\zeta; t_0-1, t_0+1)}(t)$ and $EDV^{AO}_{(\zeta; t_0-1, t_0+1)}(t)$ are given by (4.13), (4.14), (4.16), and (4.17) above (since only y_{t-1}, y_t, and y_{t+1} enter in (4.10) and (4.15)). Also, it is easy to show that as a consequence of the symmetry in the expected asymptotic diagnostics for AO models, $EDC^{AO}_{(\zeta; t_0-1, t_0+1)}(t_0) = EDC^{AO}_{(2\zeta, t_0-1)}(t_0)$ and $EDV^{AO}_{(\zeta; t_0-1, t_0+1)}(t_0) = EDV^{AO}_{(2\zeta; t_0-1)}(t_0)$. That is, the smearing effect of an AO type outlier is additive: outliers of size ζ at $t_0 - 1$ and $t_0 + 1$ are equivalent to an outlier of size 2ζ at $t_0 - 1$.

To see just how serious smearing can be in this situation, consider the case where there are outliers of size $+4$ at $t_0 - 1$ and $t_0 + 1$. Figure 5b exhibits $\dfrac{EDC^{AO}_{(+4; t_0-1, t_0+1)}(t_0)}{EDC^{AO}_{(+4; t_0-1, t_0+1)}(t_0 - 1)}$ (solid line) and $\dfrac{EDV^{AO}_{(+4; t_0-1, t_0+1)}(t_0)}{EDV^{AO}_{(+4; t_0-1, t_0+1)}(t_0 - 1)}$ (dashed line) as a function of ϕ. The expected asymptotic value of $DC(t_0)$ with outliers at $t_0 - 1$ and $t_0 + 1$ is larger than the expected asymptotic diagnostic at either outlier position for all ϕ, and has a maximum value almost six times larger! Thus, DC will be totally ineffective in revealing such a configuration of outliers. By contrast, the ratio for EDV stays below one for all ϕ, and is substantially smaller than one except for values of $|\phi|$ near one. One therefore expects DV to be far superior to DC in revealing such outlier configurations.

4.4 IO Models: AR(1) Case

The analysis of smearing for IO models parallels that for AO models. However, since the outlier occurs in the innovations of the process, the difference in the score functions for ϕ is not symmetric, as was the case for AO models. Suppose y_t is observed with an IO type outlier of magnitude ζ at t_0, i.e, ε_t is given by (3.9) with $z_t = 0$ except at $t = t_0$ where $z_t = 1$. If x_t represents the series *without* the innovations outlier, then it is easy to check that

$$y_t = \begin{cases} x_t \;, & t < t_0 \\ x_t + \zeta \phi^{t-t_0} \;, & t \geq t_0 \;. \end{cases} \tag{4.18}$$

Details concerning the calculation of the expressions to follow are provided in Appendices B and C.

Expression for $EDC^{IO}(t)$

From (4.10) we get the difference in the score functions for ϕ with and without y_{t_0}:

$$\begin{aligned} \Delta^{IO}_{(\zeta, t_0)}(t_0; \phi) = & -\frac{\phi}{1-\phi^2} - \frac{1}{\sigma^2}\Big[-\phi(x_{t_0}+\zeta)^2 \\ & + (x_{t_0}+\zeta)(x_{t_0-1} + (x_{t_0+1}+\phi\zeta)) \\ & - \frac{\phi}{(1+\phi^2)^2}(x_{t_0-1} + (x_{t_0+1}+\phi\zeta))^2 \Big] \;. \end{aligned} \tag{4.19}$$

The notation $\Delta^{IO}_{(\zeta, t_0)}(t; \phi)$ is used to indicate that we have a single IO type outlier of size ζ at time t_0. It is now straightforward to show that

$$EDC^{IO}_{(\zeta;t_0)}(t_0) = \left[\frac{\zeta}{\sigma}\right]^4 \frac{\phi^6(1-\phi^2)}{(1+\phi^2)^4} + \left[\frac{\zeta}{\sigma}\right]^2 \left[2-\phi^2\left[1+\frac{8}{(1+\phi^2)^3}\right]\right] + \frac{2(1-\phi^2)}{(1+\phi^2)^2}\,. \tag{4.20}$$

Although the dominant term is still proportional to $\left[\frac{\zeta}{\sigma}\right]^4$, it is of order $o(\phi^5)$ as $|\phi| \to 0$.

For $t < t_0 - 1$, $EDC^{IO}_{(\zeta;t_0)}(t)$ is given by (4.20) with $\zeta \equiv 0$. $EDC^{IO}_{(\zeta;t_0)}(t_0-1)$ is identical to the counterpart (4.14) for AO models (recall that (4.14) is also the value for $t = t_0 - 1$). For $t = t_0 + i$, $i = 1, 2, \cdots$,

$$EDC^{IO}_{(\zeta;t_0)}(t_0+i) = \left[\frac{\zeta}{\sigma}\right]^2 \frac{\phi^{2(i-1)}(1-\phi^2)^3}{(1+\phi^2)^2} + \frac{2(1-\phi^2)}{(1+\phi^2)^2}\,. \tag{4.21}$$

Note that the dominant term in (4.21) is proportional to $(\frac{\zeta}{\sigma})^2$ rather than $(\frac{\zeta}{\sigma})^4$. Hence, the effect of innovations outliers on EDC for $t > t_0$ is "smaller" than that at $t = t_0 - 1$ and $t = t_0$ and furthermore the effect dies out exponentially fast in t.

Expressions for $EDV^{IO}(t)$

Substituting y_t in (4.18) for x_t in (4.15) gives the difference in the score functions for σ^2:

$$\Delta^{IO}_{(\zeta,t_0)}(t_0;\sigma^2) = \frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}[-(1+\phi^2)(x_{t_0}+\zeta)^2$$

$$+2\phi(x_{t_0}+\zeta)(x_{t_0-1}+(x_{t_0+1}+\phi\zeta))$$

$$+\frac{\phi^2}{1+\phi^2}(x_{t_0-1}+(x_{t_0+1}+\phi\zeta))^2] \quad (4.22)$$

Use of (4.9) yields

$$EDV^{IO}_{(\zeta;t_0)}(t_0) = \left[\frac{\zeta}{\sigma}\right]^4 \frac{1}{2(1+\phi^2)^2} + \left[\frac{\zeta}{\sigma}\right]^2 \frac{2}{(1+\phi^2)} + 1 \quad (4.23)$$

For $t \neq t_0-1$ or t_0, $EDV^{IO}_{(\zeta;t_0)}(t)$ is given by (4.23) with $\zeta \equiv 0$, and $EDV^{IO}_{(\zeta;t_0)}(t_0-1)$ is given by (4.17), the corresponding expected diagnostic under the AO model. Unlike the AO case, $EDV^{IO}_{(\zeta;t_0)}(t_0+1)$ does not depend on ζ.

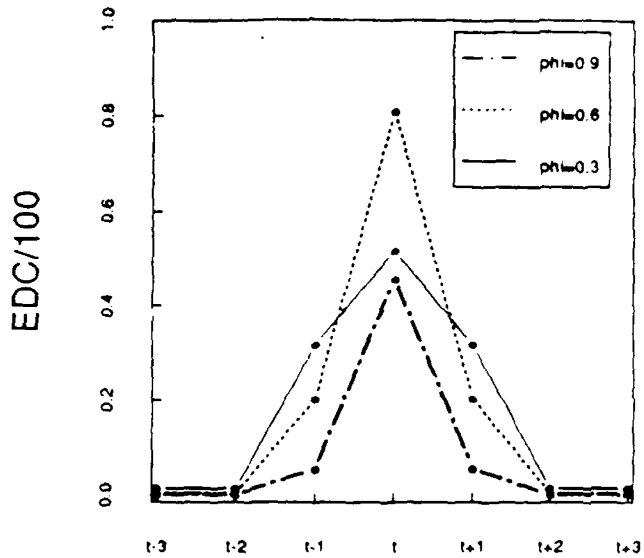
Comparisons of $EDC^{IO}(t)$ and $EDV^{IO}(t)$

With IO models, the behavior of the smearing effects of an outlier at t_0 differ even more dramatically for *EDC* and *EDV*. For *EDC*, the effect of smearing is not restricted to points immediately adjacent to the time of occurrence t_0 of an outlier. Specifically, an outlier affects $EDC^{IO}_{(\zeta;t_0)}(t)$ for all $t \geq t_0-1$ (i.e., leaving out the previous point or any future point). By way of contrast, the effects of an outlier at t_0 are seen only at t_0-1 and t_0 for *EDV*!

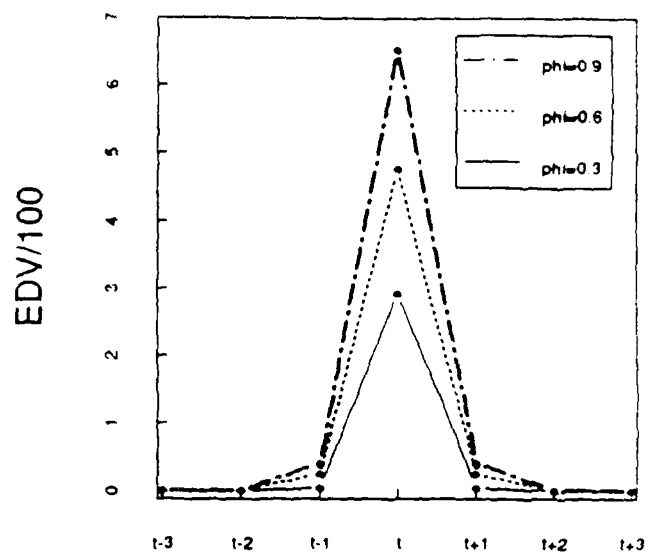
Figures 4c and 4d display $EDC^{IO}_{(+4;t_0)}(t)/100$ and $EDV^{IO}_{(+4;t_0)}(t)/100$ for $t=t_0-3,\ t_0-2,\ \cdots,\ t_0+3$ with $\phi=.3,\ .6$, and $.9$. The severe smearing of *DC* at $t=t_0-1$ is reflected in Figure 4c, where $EDC^{IO}_{(+4,t_0)}(t_0-1)$ dominates $EDC^{IO}_{(+4;t)}(t_0)$. This is also demonstrated by Figure 5c, which shows $\dfrac{EDC^{IO}_{(+4;t_0)}(t_0-1)}{EDC^{IO}_{(+4;t_0)}(t_0)}$ (solid line) and $\dfrac{EDV^{IO}_{(+4;t_0)}(t_0-1)}{EDV^{IO}_{(+4;t_0)}(t_0)}$ (dashed line). The ratio

for EDC^{IO} is greater than one for most values of ϕ, while the ratio for EDV^{IO} stays well below unity, except for $|\phi|$ close to 1.

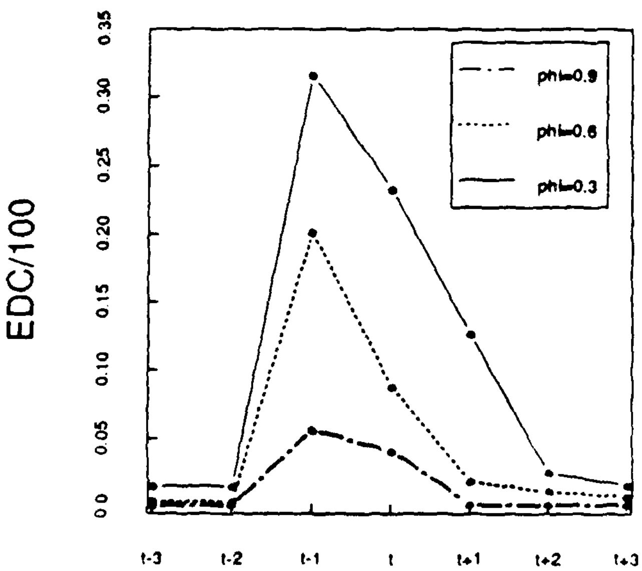
These results extend to AR(p) models: dominant values of EDC^{IO} can occur at the p consecutive times *preceeding* an isolated outlier. The use of $EDV(t)$ is obviously preferred for IO (as well as AO) situations.



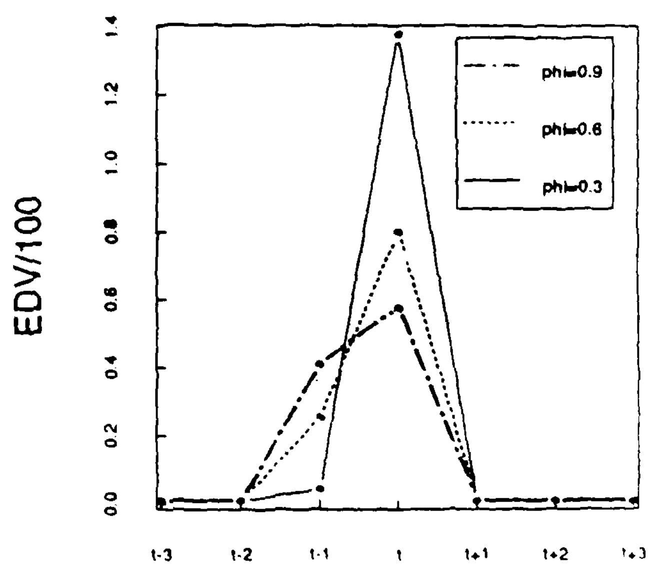
a) EDC: AO type outlier of size +4 at time t



b) EDV: AO type outlier of size +4 at time t



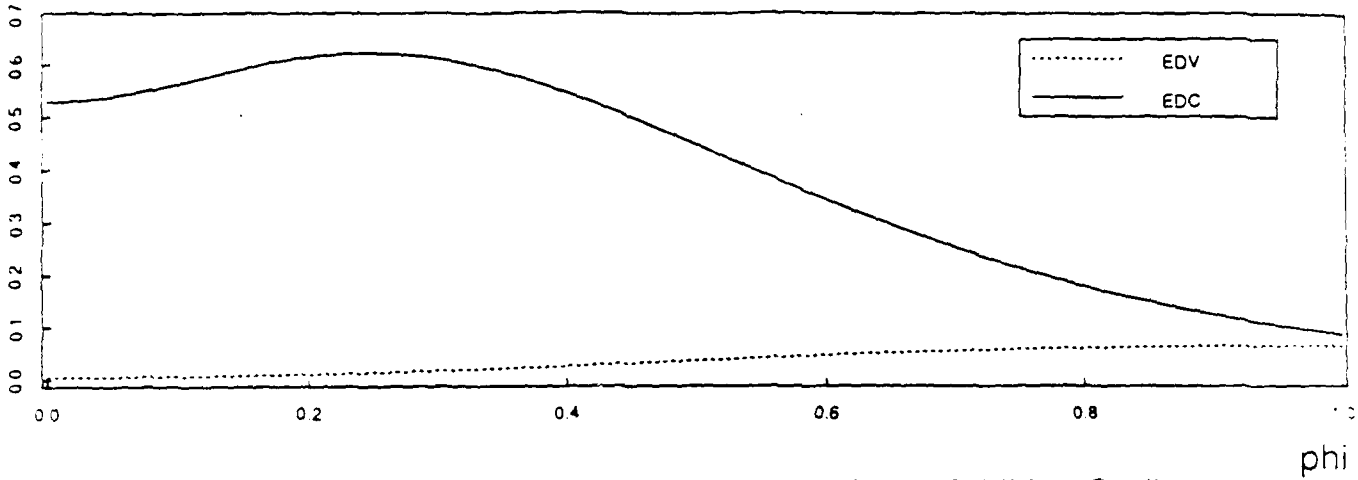
c) EDC: IO type outlier of size +4 at time t



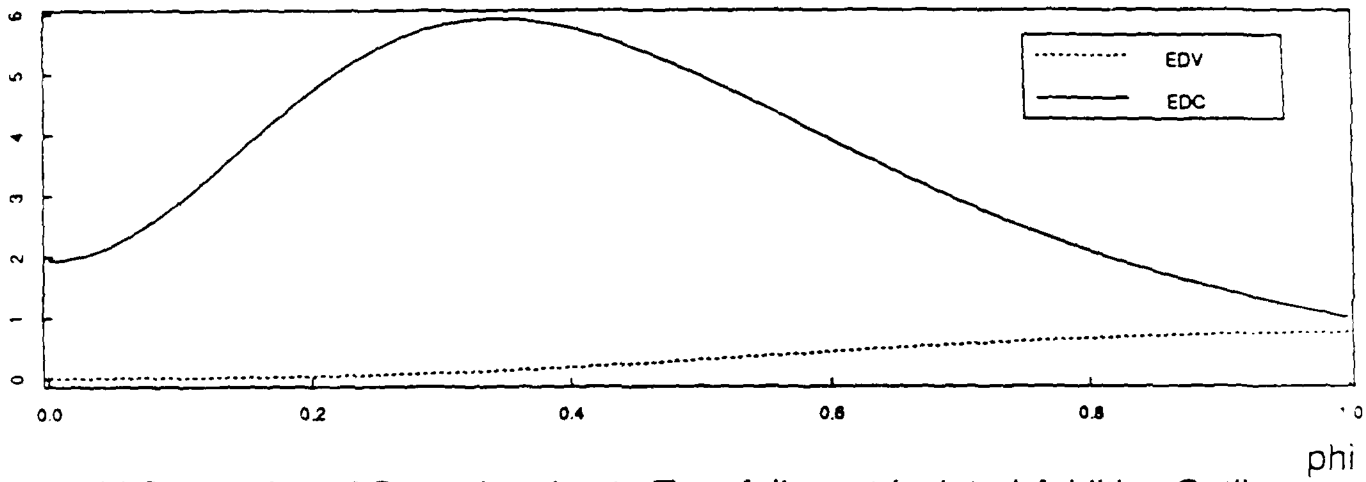
d) EDV: IO type outlier of size +4 at time t

Expected Asymptotic Diagnostic Curves

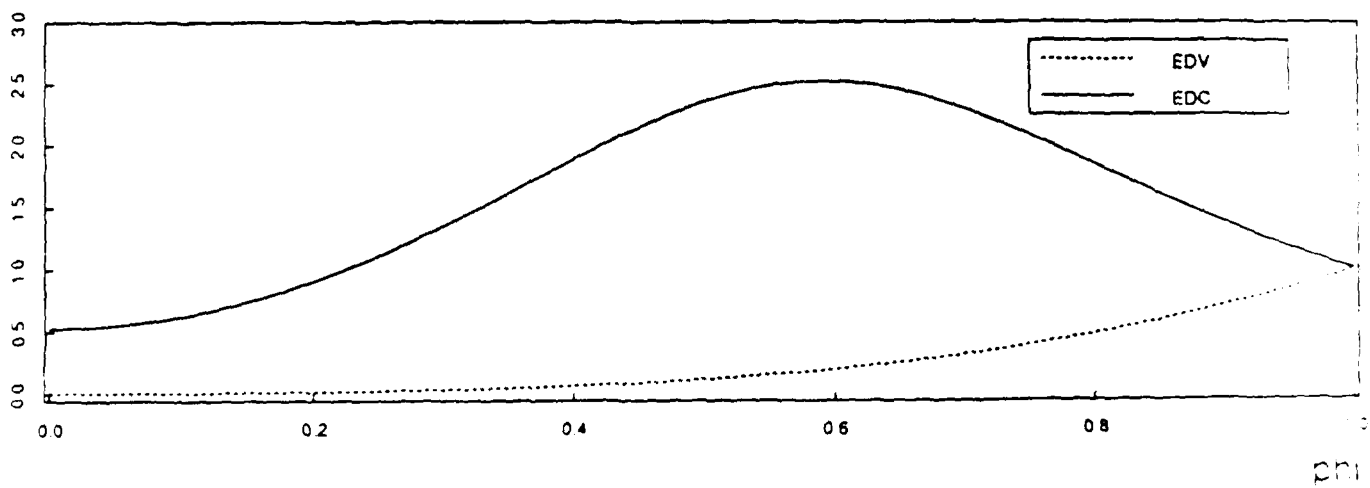
Figure 4



a) Proportion of Smearing due to an Isolated Additive Outlier



b) Proportion of Smearing due to Two Adjacent Isolated Additive Outliers



c) Proportion of Smearing due to an Isolated Innovations Outliers

Expected Asytmptotic Diagnostics: Smearing Plot

Figure 5

5. Diagnosing IO Versus AO

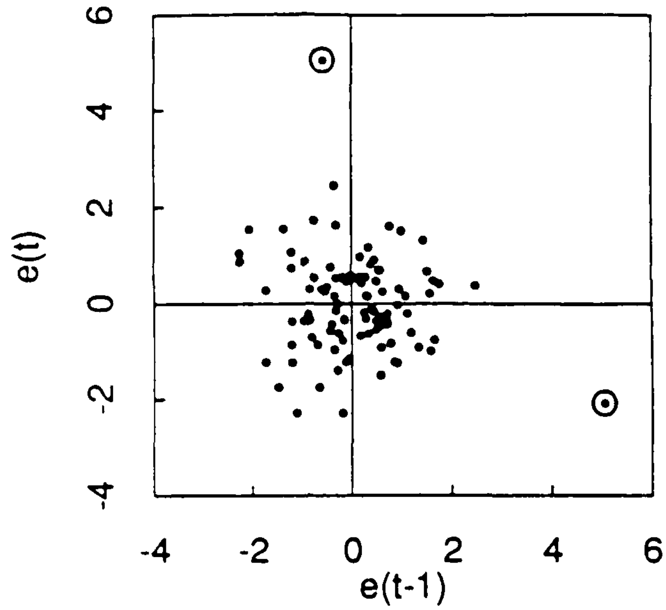
Notice that Figures 1c and 2c appear to be very much alike, in spite of the fact that the outlier at $t = 28$ is an AO in Figure 1b and an IO in Figure 2b. DV identifies the outlier in both cases, but provides no additional information as to the outlier type. On the other hand, the distinctively different behavior of DC for the two cases allows one to use the auxiliary information contained in DC to help decide whether an outlier is IO or AO: If DC is smeared to both sides of the time of occurrence of the outlier as identified by DV, then the outlier is probably AO. If DC is large prior to the outlier time identified by DV, but small at the outlier time, then the outlier is probably IO.

A more formal way of determining whether an outlier identified by DV is IO or AO is to use a robustified version of Fox's test (1972), as described in Martin and Zeh (1977). See also the non-robust use of Fox type tests in an outlier identification and model fitting scheme proposed in Hillmer, Bell and Tiao (1983).

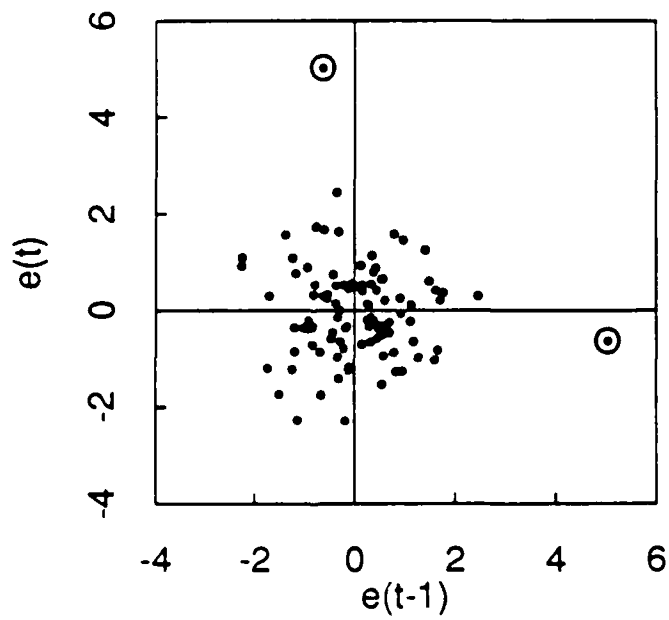
A less formal way of determining whether an outlier is IO or AO is to examine a lag-1 scatter plot of the residuals. As was pointed out by Martin and Zeh (1977), IO's tend to fall near the abscissa and ordinate of such a plot, whereas AO's tend to appear away from the abscissa and ordinate, assuming that robust parameter estimates have been used to form the residuals. In the present context, we recommend using the parameter estimates obtained with the outliers identified by DV deleted.

Figures 6a and 6b display the lag-1 residuals plot for the data of Figures 1a and 2a respectively. The circled points are (e_{27}, e_{28}) and (e_{28}, e_{29}), where e_t is the one-step ahead prediction residual for time t. As expected, the (e_{28}, e_{29}) point for AO falls further from the abscissa than in the IO case. In particular, for the IO case, the ordinate value is well within the bulk of the data, while in the AO case, it is near the extreme lower range of the data (it is the third smallest value).

Combining the DC diagnostics with the lag-1 scatterplots, we obtain a convincing graphical display identifying the type of an isolated outlier. However, when a patch of outliers is present, these techniques do not directly apply, and determination of the outlier type is a more subtle problem.



a) Example 3.1 (AO): Lag 1 Scatterplot of Residuals



b) Example 3.2 (IO): Lag 1 Scatterplot of Residuals

Examples 3.1 and 3.2 (continued): Identifying AO vs IO

Figure 6

6. Overall Strategy

In this section we present an overall strategy for ARIMA model fitting using leave-k-out diagnostics. In Section 6.1, to handle problems caused by "masking", we imbed the approach of Section 3 for determining the length of a patch of outliers in an iterative deletion procedure. We discuss more flexible subset deletion techniques in Section 6.2 to handle cases where the iterative deletion procedure fails. Finally, an overall strategy for model identification and fitting is given in Section 6.3.

6.1 An Iterative Deletion Strategy

The masking of influential points (e.g., outliers) by other influential points is a problem encountered in all types of diagnostics. As we have already seen, masking caused by a single patch of outliers can be handled adequately by leave-k-out diagnostics. However, sometimes the presence of a gross outlier will have sufficient influence so that deletion of aberrant values elsewhere in the series has little effect on the estimate. More subtle types of masking occur when moderate outliers occur in close proximity to one another. These types of masking can often be effectively uncovered by an iterative deletion process which consists of removing suspected outliers from the data, and recomputing the diagnostics.

To deal with problems caused by masking, we build upon the initial patch length determination strategy of Section 3 as follows:

Step 1

Run leave-k-out diagnostics on the data, for $k = 1, 2, \ldots,$ until either: (a) the length of the most influential (significant) patch is determined using the guidelines of Section 3, or (b) $k = K_{max}$, where K_{max} is determined by the user. In principle, K_{max} is the length of the longest patch of outliers thought to be present in the data. However, computational costs require that K_{max} is reasonably small (see the run time results in Section 8.2; for "short" time series, i.e., $n < 250$, se

$K_{max} = 5$ will often reveal most if not all problems with the data). Case b can result from two possibilities: either no influential observations were detected, or the length of an influential patch is ill-determined (according to the guidelines of Section 3). In the latter case, which may be due to the presence of a patch of length greater than K_{max}, we determine the "most influential" patch as that corresponding to the most significant diagnostic.

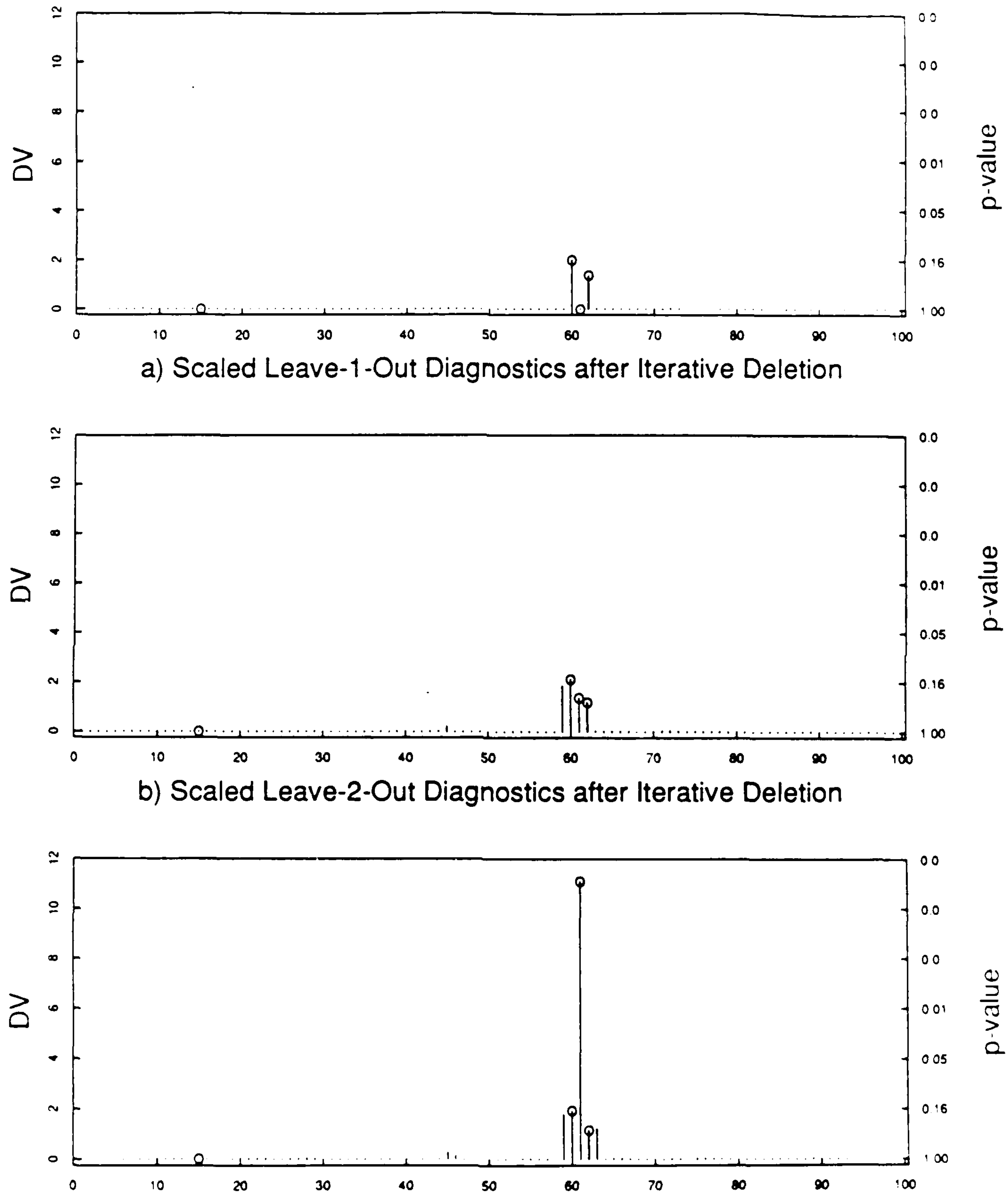
Step 2

If no influential points are found, then conclude the analysis. If influential points are found, then delete the most influential points as identified in step 1, and go back to step 1. The new leave-k-out coefficients should be scaled according to the MLE computed with the outliers removed, so as to gauge additional influence of the remaining points!

The next two artificial examples illustrate the efficacy of the iterative deletion procedure in handling problems caused by masking.

Example 3.3 (continued):

With the above modified guidelines, we would identify y_{15} as an outlier after running leave-1-out and leave-2-out diagnostics (see Figures 3b and 3c). Performing iterative deletion, we "remove" y_{15} from the data (i.e., treat y_{15} as missing) and recompute leave-k-out diagnostics for $k = 1, 2, 3$. These are displayed for DV in Figures 7a–7c, and give convincing evidence of a further patch of 3 outliers centered at $t = 61$. Note that the pattern of diagnostics after iterative deletion is nearly identical (except for values associated with y_{15}) to the original set of diagnostics (cf. Figures 3b–3d). However, the magnitude of the diagnostics is much larger after iterative deletion! This is quite typical: a non-adjacent outlier(s) will mask other outliers by decreasing the magnitude of the diagnostics, but not altering the pattern.



a) Scaled Leave-1-Out Diagnostics after Iterative Deletion

b) Scaled Leave-2-Out Diagnostics after Iterative Deletion

c) Scaled Leave-3-Out Diagnostics after Iterative Deletion

Example 3.3 (continued): Simulated MA(-.5) With 1 Patch and 1 Isolated Outlier

Figure 7

Example 6.1: Simulated MA(1), $\theta = -.5$, AO Model with 1 Patch and 1 Adjacent Isolated Outlier

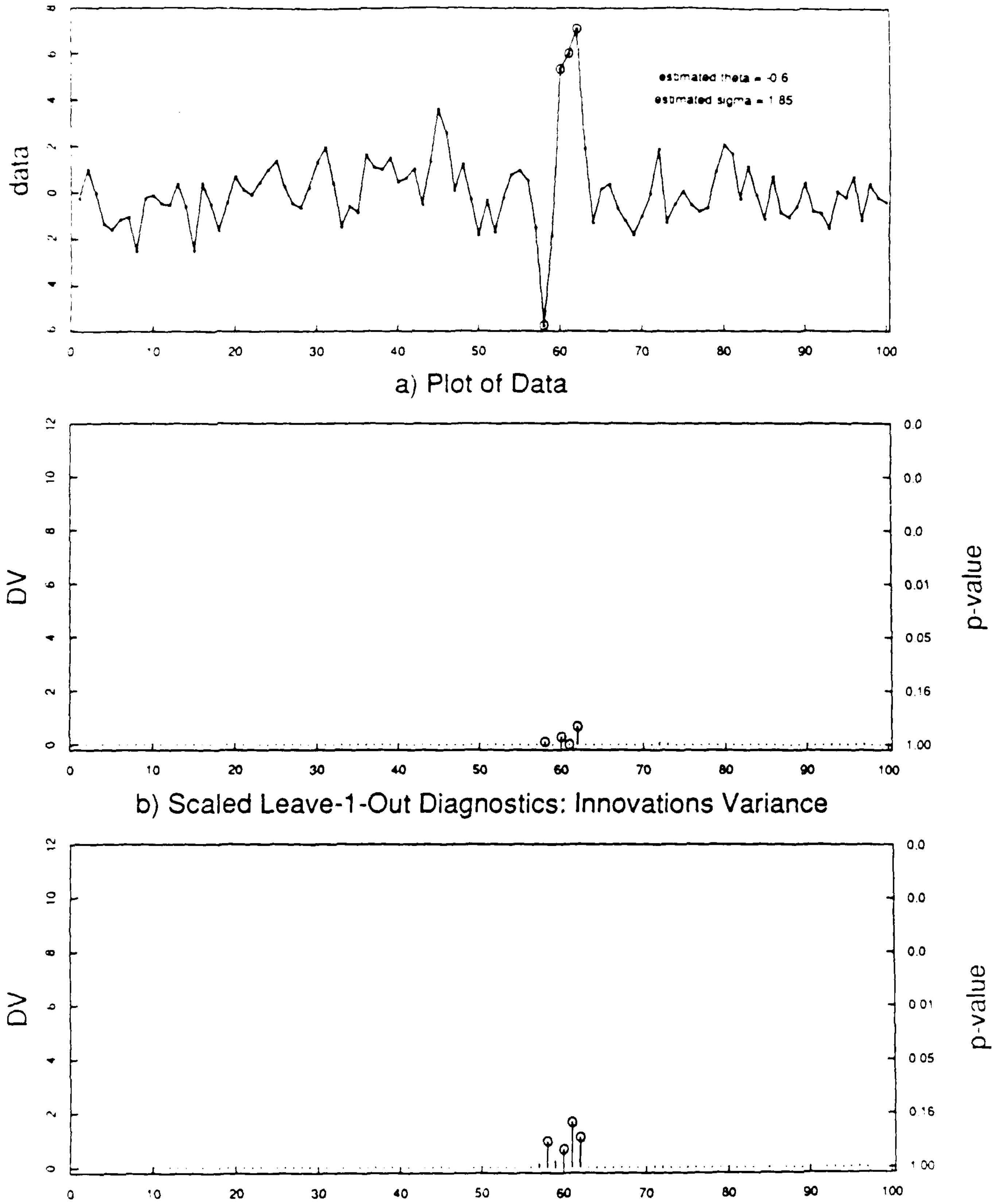
The data is the same as that used in Example 3.3, except that the isolated outlier is moved from point 15 to point 58, adjacent to the patch of outliers at points 60–62. The data is plotted in Figure 8a. Leave-1-out, leave-2-out, leave-3-out and leave-4-out diagnostics for DV are given in Figures 8b–8e.

The masking is much more severe than in Example 3.3: the isolated outlier is now completely masked for the leave-1-out case (cf. Example 3.3), though the patch still shows up prominently in the leave-3-out diagnostics. However, leave-4-out diagnostics are only slightly more significant than leave-3-out, and the pattern of smearing is consistent with a patch of 3 outliers. So following our strategy, we delete points 60–62 and recompute the diagnostics. The isolated outlier is now easily identified by the recomputed leave-1-out diagnostics of Figure 8f: removal of the patch eliminates the masking problem.

6.2 Local Stuctural Changes and Flexible Subset Deletion Techniques

Until now, we have concentrated on influential points in the form of outliers. However, influential points may also be due to other types of disturbances, such as level shifts or variance changes. The iterative deletion procedure of Section 6.1 is often effective for uncovering these types of problems. However, the procedure will sometimes fail in the presence of an influential patch longer than K_{max}; an example of this is provided in Section 7. Masking may prevent a long patch, or any points in the patch, from being detected. Even when the patch is detected, if the disturbance spans a time period considerably greater than K_{max}, the iterative deletion may require an intolerable number of iterations. To handle these failures, we adopt a data oriented "free and easy" approach to flexible subset deletion.

We consider deviating from the iterative deletion procedure and using the flexible approach primarily in two kinds of situations. First, when examining the data and the



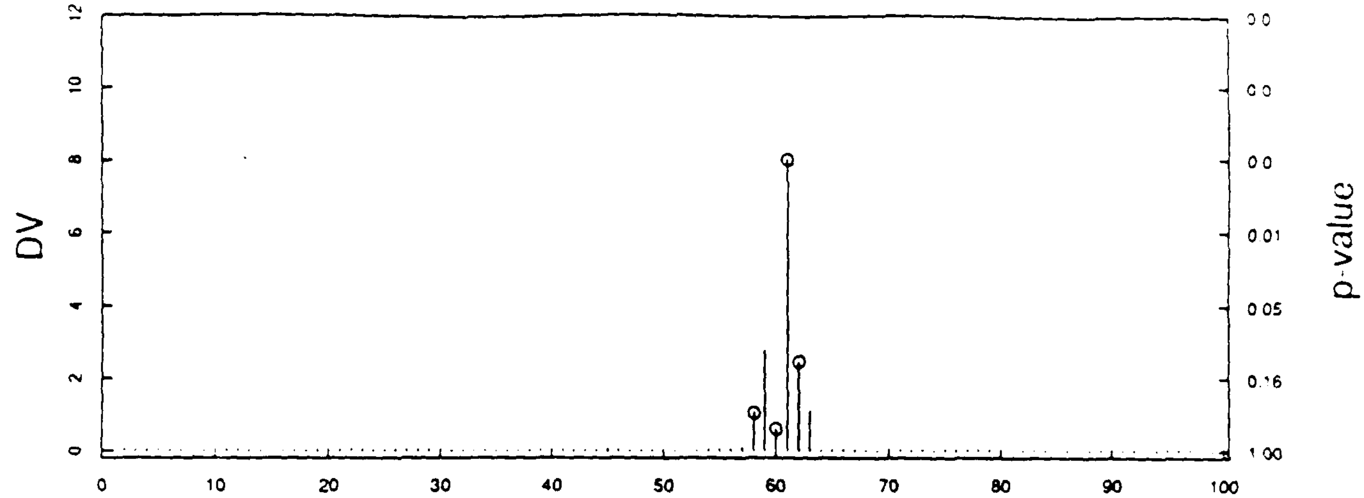
a) Plot of Data

b) Scaled Leave-1-Out Diagnostics: Innovations Variance

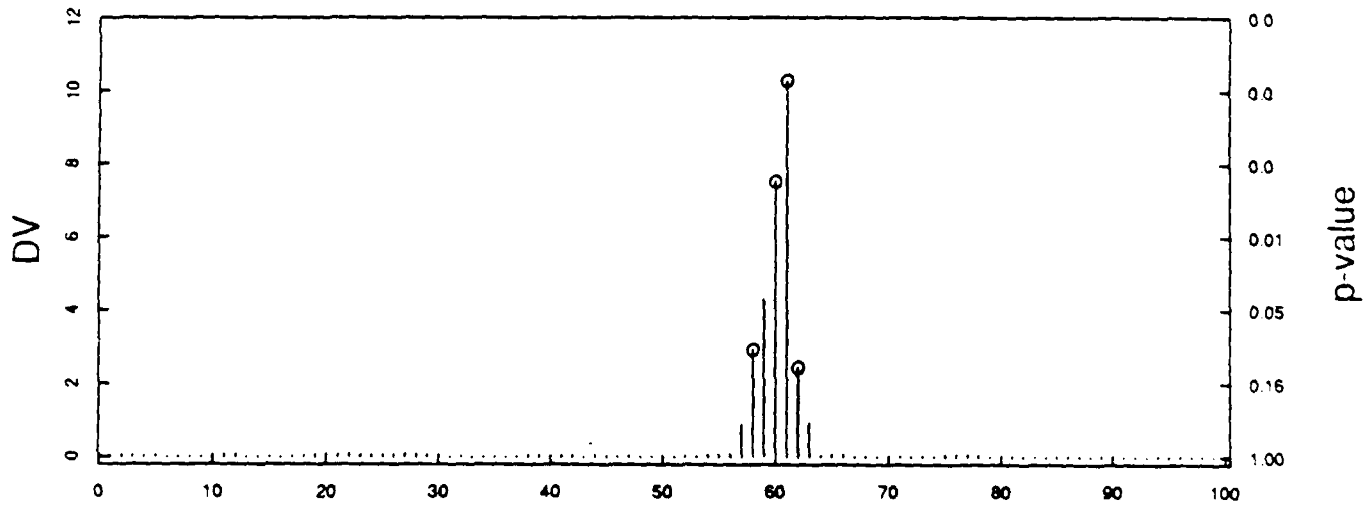
c) Scaled Leave-2-Out Diagnostics: Innovations Variance

Example 6.1: Simulated MA(-.5) With 1 Patch and 1 Adjacent Isolated Additive Outlier

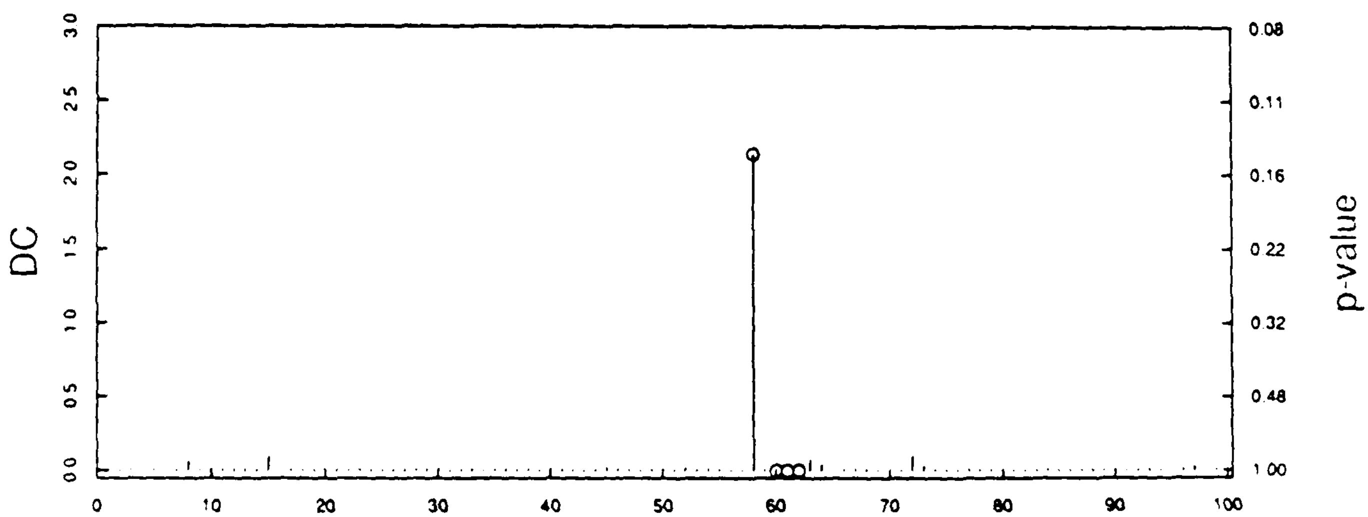
Figure 8



d) Scaled Leave-3-Out Diagnostics: Innovations Variance



e) Scaled Leave-4-Out Diagnostics: Innovations Variance



f) Scaled Leave-1-Out Diagnostics After Iterative Deletion

Example 6.1: Simulated MA(-.5) With 1 Patch and 1 Adjacent Isolated Additive Ou[illegible]er

Figure 8

residuals, the analyst may suspect a structural change in the data. Second, the leave-k-out diagnostics may indicate at a local disturbance of duration greater than K_{max}, (e.g., as when the patch length is ill-determined; see step 1 of Section 6.1). In either case, flexible subset deletion techniques can help identify the structure more precisely.

An attractive way of carrying out flexible subset deletions is through the use of interactive graphics on a computer workstation. Candidate subsets are identified on computer graphics plots of the data and/or the residuals, and DV is computed for such subsets. For example, if the analyst believes a local level shift is present somewhere between the times t_0 and t_1, then he/she would compute DV for a judicious selection of patches between t_0 and t_1 in order to clarify the jump points. This procedure may easily be carried out with the aid of a mouse and appropriate software (see Section 8.2).

A non-interactive and computationally expensive approach is to run leave-k-out diagnostics on the data for selected values of k between K_{max} and $n/2$. A ''top down'' approach for selecting k is to use $k = [n/2], [n/4], \ldots, [n/2^r]$ where r is the largest integer such that $n/2^r > K_{max}$. An alternative ''bottom up'' approach is to choose $k = 2^s, 2^{s+1}, \ldots, 2^t$ where s is the smallest integer such that $2^s \geq K_{max}$ and t is the largest integer such that $2^t \leq n/2$. From these diagnostics, the disturbances can often be clarified.

Another application of the ''bottom up'' or ''top down'' diagnostics is as a final check on the model: see Section 6.3.

6.3 Model Identification and Overall Strategy

The foregoing analysis presumed that the degree of differencing and the order for the model was known. In practice, this is rarely the case, and the model must be determined by some criteria such as the Box Jenkins identification procedure. However, outliers may cause improper model specification. To handle order selection in the presence of outliers and structural changes, we embed the iterative deletion strategy in an iterative procedure sim...

to that used by Tsay (1986).

Overall Strategy

Step 0: Tentative Identification

Using the Box Jenkins methodology, determine a tentative model. This involves specifying the degree of differencing and selecting the order of the ordinary and seasonal ARMA components.

Step I: Iterative Deletion

Perform leave-k-out diagnostics using the iterative deletion strategy of Section 6.1 until: no additional influential points are detected, or a model change is suspected. Recall that the second situation may be triggered in two ways, as described in Section 6.2. In the first case, go to step II, while in the second case, go to step IV.

Step II: Confirming Model Order

With the observations identified as influential removed, i.e., treated as missing data, determine the order of the model once again. If the same model is selected, then go to step III. Otherwise, remove the influential observations and go back to step I.

Step III: Final Check

To ensure a longer patch is not missed, perform the "bottom up" or "top down" approach as described in Section 6.2. If nothing is revealed, then conclude the analysis in the usual way. On the other hand, if a structural change is detected, then go to step IV.

Step IV: Handling Structural Changes

Split the data according to the conjectured model change point(s). Analyze separately those parts which are sufficiently long. That is, go to step 0, treating each sufficiently long section of data as different series, and ignoring any segments which are longer than K_{max}, but not long enough to warrant model fitting.

Analysis of Residuals and Influential Points

In the presence of outliers and structural changes, the usual prediction residuals are often misleading for identifying the influential observations. Instead, we recommend examining the residuals based on the predictions formed when the observations identified as influential are treated as missing. Since the predictions are not distorted by influential observations, this procedure reveals outliers and structural changes more clearly.

After selecting the ''final'' model, a careful analysis of the influential data points should be carried out. Of particular interest is the determination of any physical causes or events related to such points. Also, one may be able to categorize influential points as isolated or patches of outliers, or perhaps associate them with a level shift or variance change. Points diagnosed as an outliers can be further classified by type (AO versus IO) using the techniques described in Section 5.

Use of Intervention Analysis

A variety of structural changes, such as outlier patches, level shifts, and even variance shifts, which may be detected by the leave-k-out strategy, can be handled by intervention analysis, as in Box and Tiao (1975). The prediction residuals for local structural changes provide information which may suggest a small palette of intervention "shapes". We note that the diagnostics may suggest intervention analysis which might be otherwise overlooked because the investigator was unaware of any particular "cause" (e.g., policy change).

7. Applications to Real Data

In this section, we analyze two economic time series using the strategy articulated in Section 6. The first series is relatively well behaved, except for several patches of outliers. The second is more difficult to model, since it contains several local nonstationarities and disturbances, including level shifts and a variance change. For brevity, we omit the details of model selection in the examples to follow, and concentrate instead on the diagnostics.

Example 7.1: Exports to Latin American Republics: 1966–1983

In this example, we study monthly unadjusted data on exports from the United States to Latin American Republics. This series was examined by Burman (1985), who focused on outliers and forecasting in U.S. Census Bureau data. A plot of the logarithm of the data is given in Figure 9a; the circled values represent points eventually deleted from the series. Following Burman, we fit an $IMA(0,1,2)$ to this series, and the residuals from the MLE fit are plotted in Figure 9b.

Leave-k-out diagnostics for $k = 1,2,3$ are displayed in Figures 9d–9f for DV. Again, for clarity, plots for DC are omitted in this example. Leaving out longer patches reveals nothing new, since the series is dominated by the outliers at 1/69 and 2/69 (i.e., Jan. 1969 and Feb. 1969). The effect on the innovations variance of leaving these two points out is dramatic (p-value < .0001): see Figure 9e. It is unlikely that a broader patch of outliers is present in this time period, since leave-3-out diagnostics yield no increased significance, and the smearing is consistent with a patch of two outliers. Note that the plots also hint at outliers in the last quarter of 1971. In fact, $DV(\cdot,2)$ is clearly significant for other points (e.g., 10/71) and is "masked" since the scale of the diagnostic at 1/69 and 2/69 is so large.

Following the strategy of Section 6, we remove the points at 1/69 and 2/69 and recompute the diagnostics for $k = 1,2,3,4$. The results of the first round of iterative deletion are displayed in Figures 9g–9j. Using the guidelines of Section 3, DV identifies the

patch 9/71, 10/71, and 11/71 as outliers, with a p-value $< .01$. Evidence for including 9/71 as part of the patch is weaker than for 10/71 and 11/71: the increase of $DV(3,t)$ over $DV(2,t)$ for $t = 10/71$ is relatively small. However, the pattern of smearing in Figures 9h-9j is more consistent with a patch of three outliers than with a patch of two. Hence, these points are removed, and the diagnostics are recomputed.

Leave-k-out diagnostics for $k = 1, 2, 3, 4$ for the second round of iterative deletion are displayed in Figures 9k–9n. These plots are noisier, but an influential patch at 12/76, 1/77, and 2/77 is clearly identified. The values for DV warrant deletion of these points (p-value $< .30$), though the "significance" is much smaller than in previous rounds.

A third round of iterative deletion was performed. The resulting leave-k-out diagnostics for $k = 1, 2, 3$ are displayed in Figures 9o–9q. Again, following the guidelines of Section 3, a patch of outliers at 1/78–2/78 is identified, though just barely (p-value $\approx .45$). Note that leave-1-out diagnostics do not pick up the patch: we need leave-2-out to identify these points as influential.

Other potential outliers are indicated by Figure 9o (10/68 and 10/70), but are associated with fairly high p-values. These points correspond to moderately large residuals (see Figures 9b and 9c), but evidently do not significantly influence the estimates of the parameters. Running another round of iterative deletion (not shown) yields little change in the significance of these points or any others.

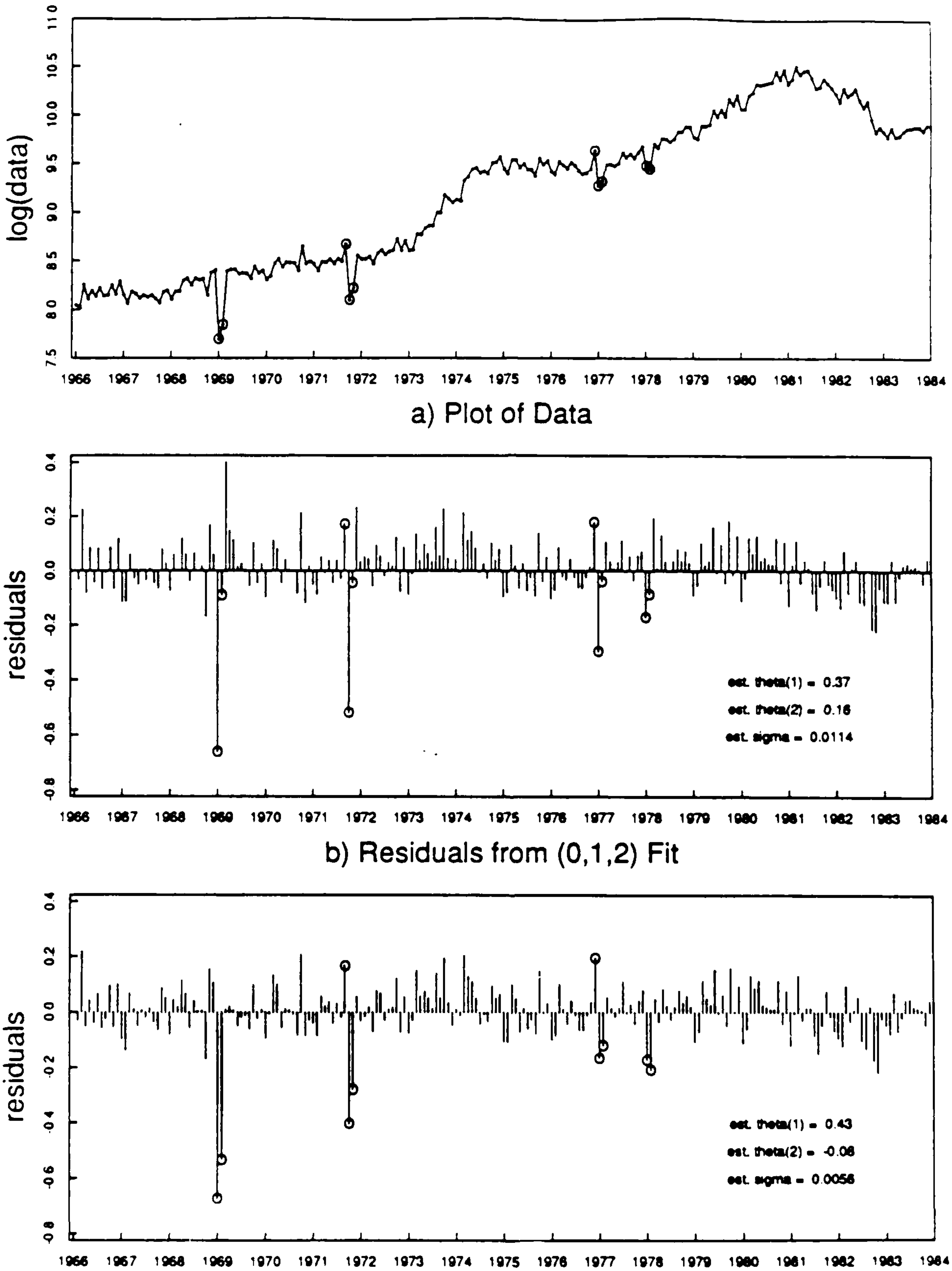
In the final analysis, four groups of outliers were identified and removed using three rounds of iterative deletion. The points which were deleted at each stage, and the corresponding MLE's, are given in Table 1. Removal of the influential points results in a drop in the estimated innovations variance by an impressive factor of two. The first two groups of outliers (1/69–2/69 and 9/71–11/71) correspond to dock strikes and forestalling yielding large negative and positive outliers respectively. The other groups (12/76–2/77 and

1/78–2/78) have no known cause, and exert considerably less influence on the estimated parameters. Burman (1985) identified the first two groups as outliers, along with 10/68 and 10/70, using the model based methodology of Hillmer, Bell, and Tiao (1983). Note that the latter two points show up in the leave-k-out diagnostics (see Figure 9o), but not so prominently as the other patches at 12/76–2/77 and 1/78–2/78, which were not identified by Burman. Once again, we see the importance of searching for influential patches as well as isolated outliers.

Table 1:
Parameters Fit to Export Data

Iteration Step	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\sigma}^2$	Points Deleted
0	.367	.160	.0114	—
1	.460	.003	.0083	1/69, 2/69
2	.448	−.041	.0066	9/71, 10/71, 11/71
3	.431	−.058	.0060	12/76, 1/77, 2/77
4	.43	−.08	.0056	1/78, 2/78

The residuals based on one-step predictions computed from the data with the outliers removed (i.e., treated as missing data) are given in Figure 9c. To obtain the predicted values, the MLE estimated with the outliers removed was used. The general pattern is similar to the original set of residuals (see Figure 9b), but with an important difference: the large residuals in Figure 9c correspond to the points identified as outliers in the above analysis. Specifically, that last outlier in each patch, masked in Figure 9b, shows up prominently in the residual plot of Figure 9c. Correspondingly, the residuals following the patch of outliers, which are large in Figure 9b, reveal nothing unusual in Figure 9c. Thus, the plot of residuals



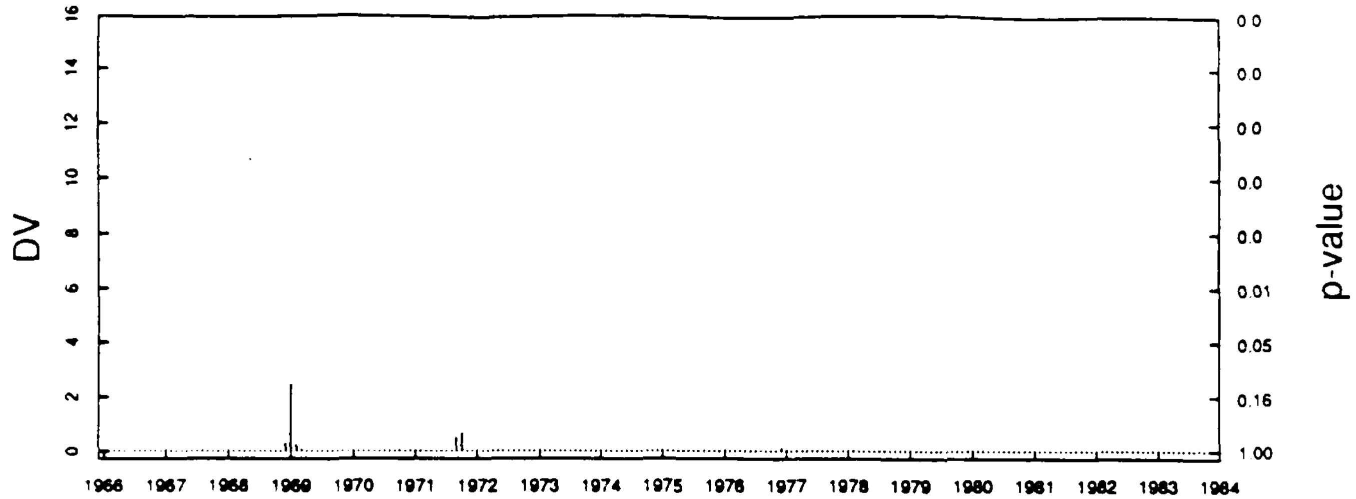
a) Plot of Data

b) Residuals from (0,1,2) Fit

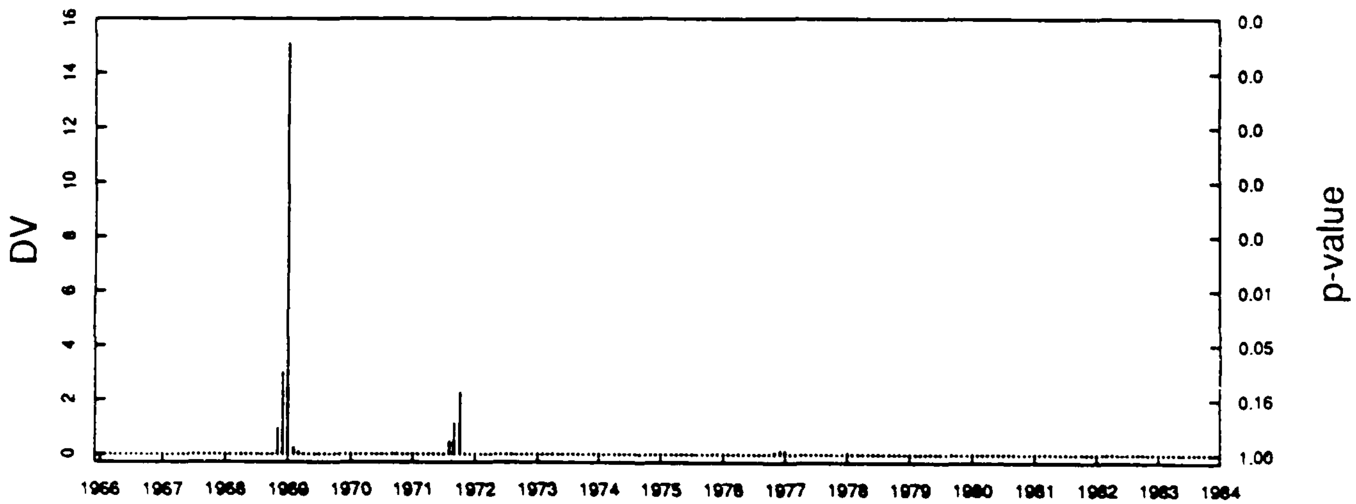
c) Residuals from (0,1,2) Fit: Outliers Removed

Example 7.1: Log of Exports to Latin America

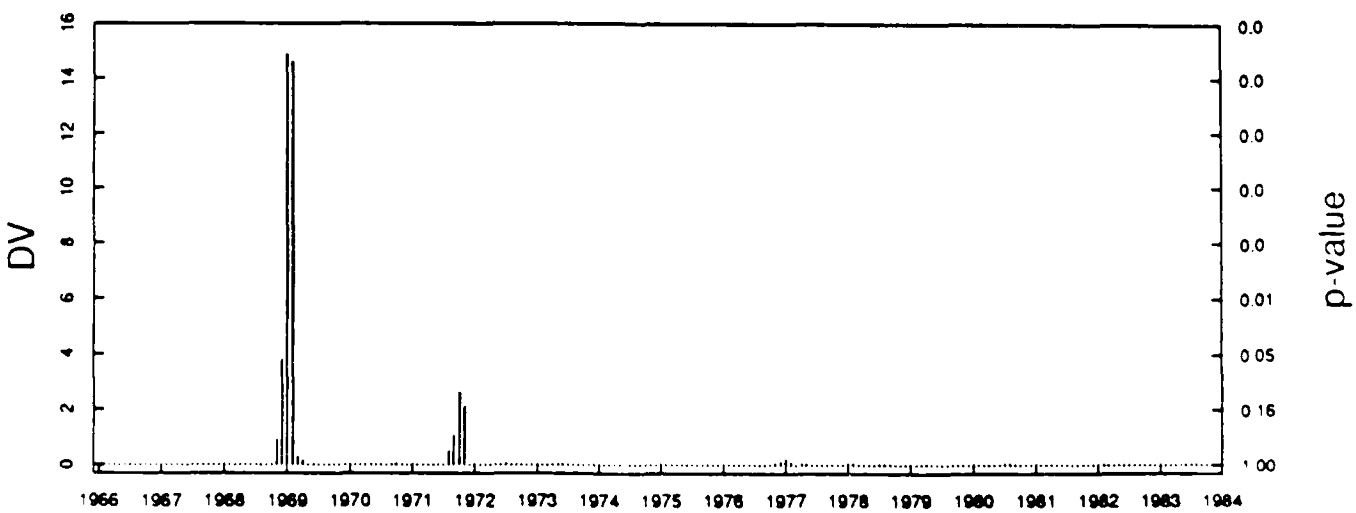
Figure 9



d) Scaled Leave-1-Out Diagnostics: Innovations Variance



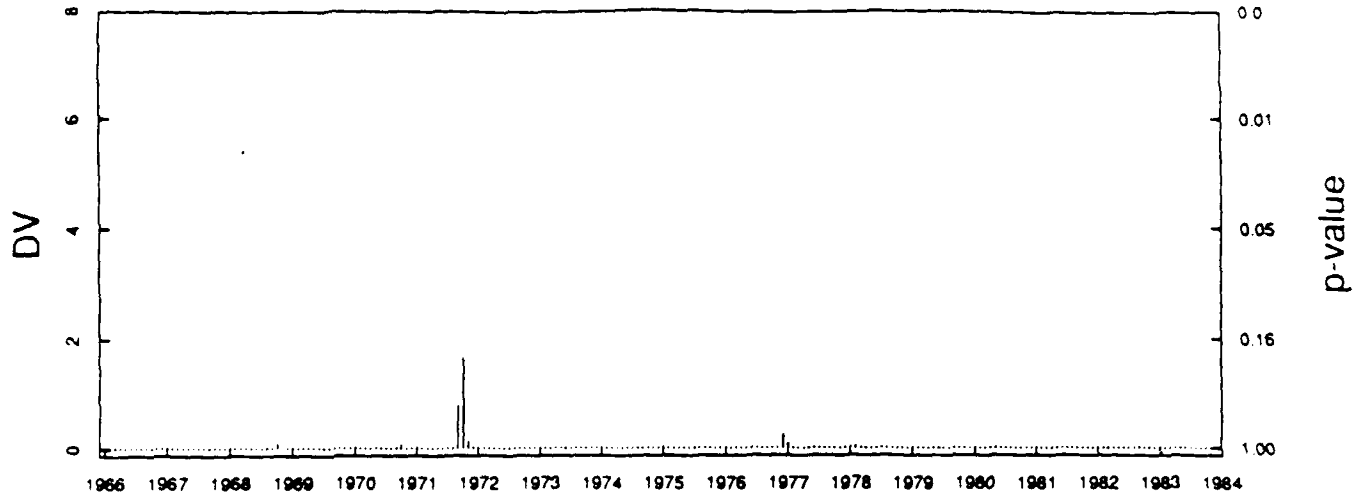
e) Scaled Leave-2-Out Diagnostics: Innovations Variance



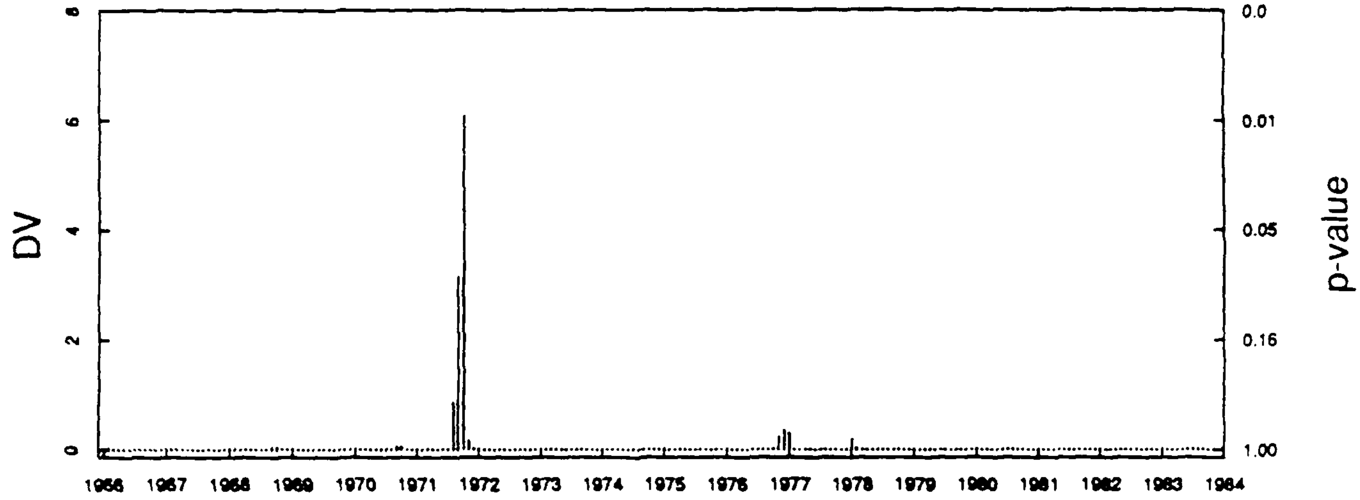
f) Scaled Leave-3-Out Diagnostics: Innovations Variance

Example 7.1: Log of Exports to Latin America

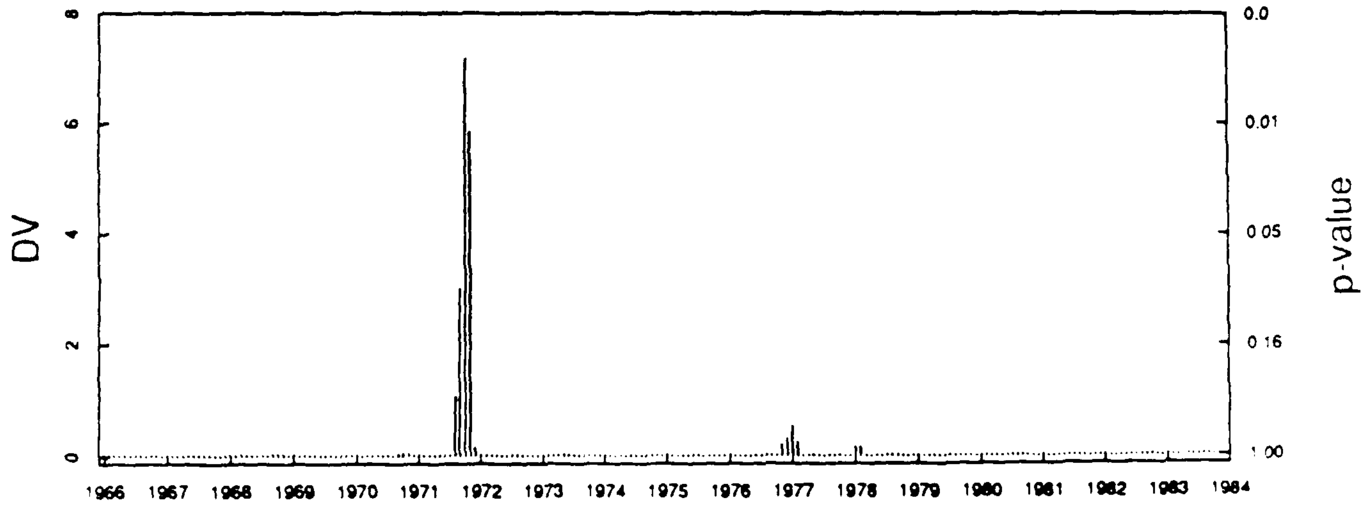
Figure 9



g) DV: Leave-1-Out after One Round of Iterative Deletion



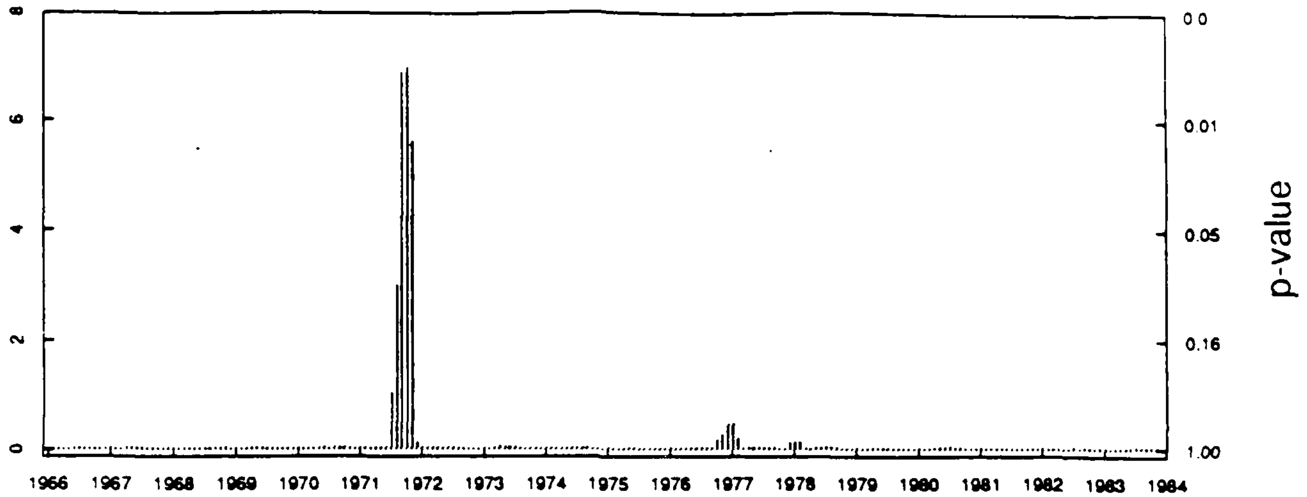
h) DV: Leave-2-Out after One Round of Iterative Deletion



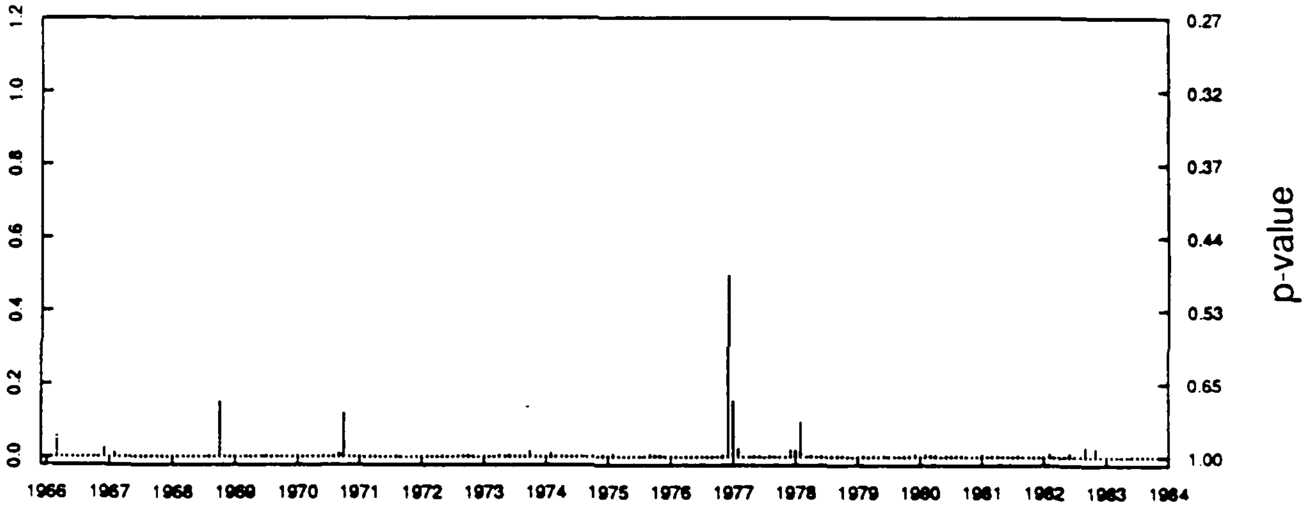
i) DV: Leave-3-Out after One Round of Iterative Deletion

Example 7.1: Log of Exports to Latin America

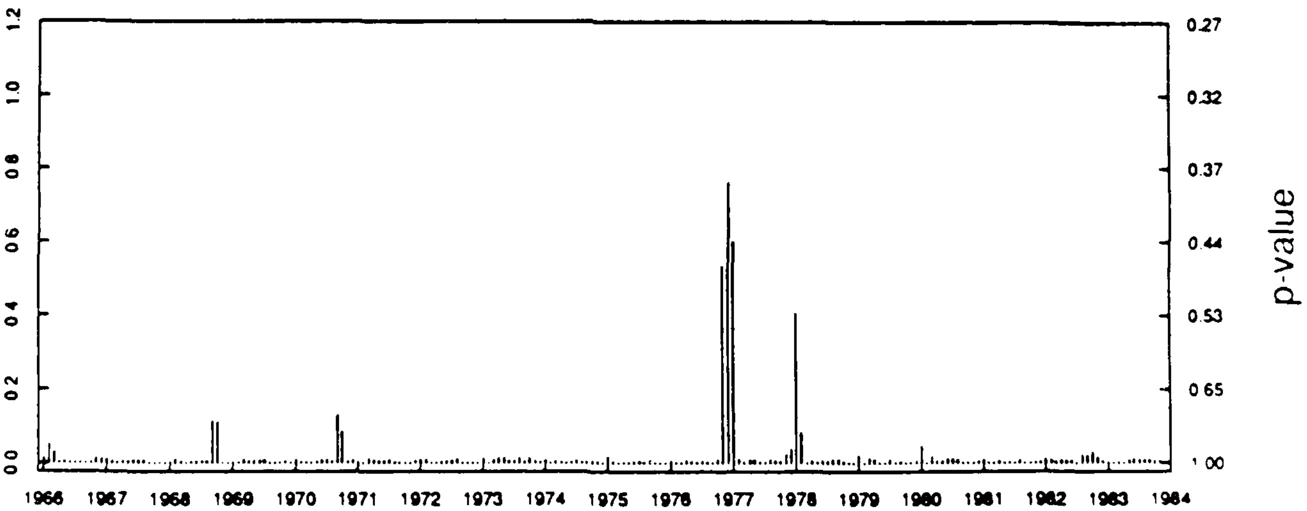
Figure 9



j) DV: Leave-4-Out after One Round of Iterative Deletion



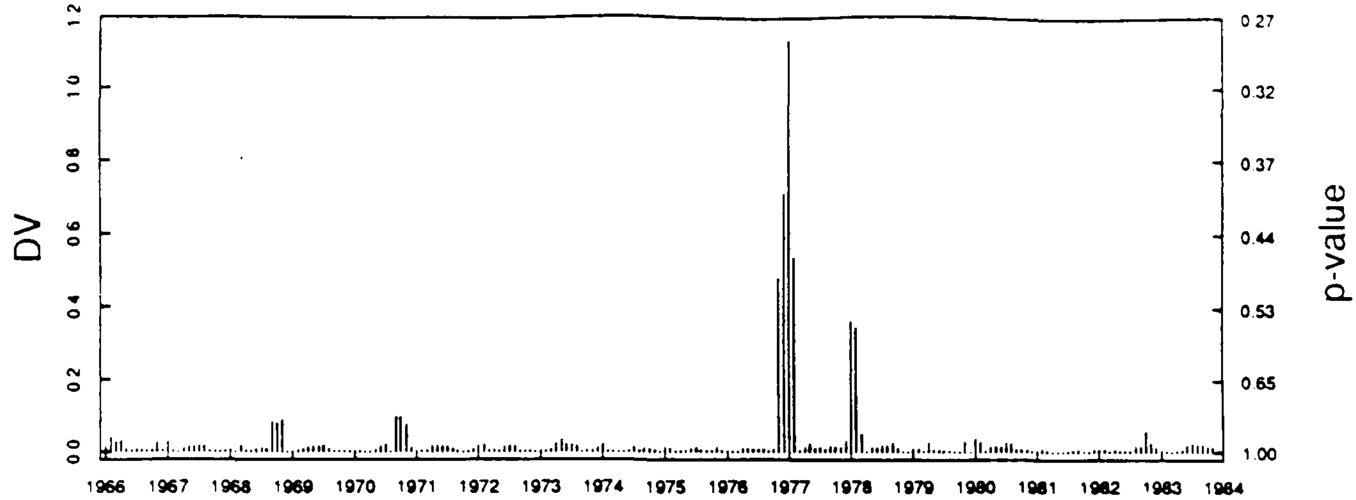
k) DV: Leave-1-Out after Two Rounds of Iterative Deletion



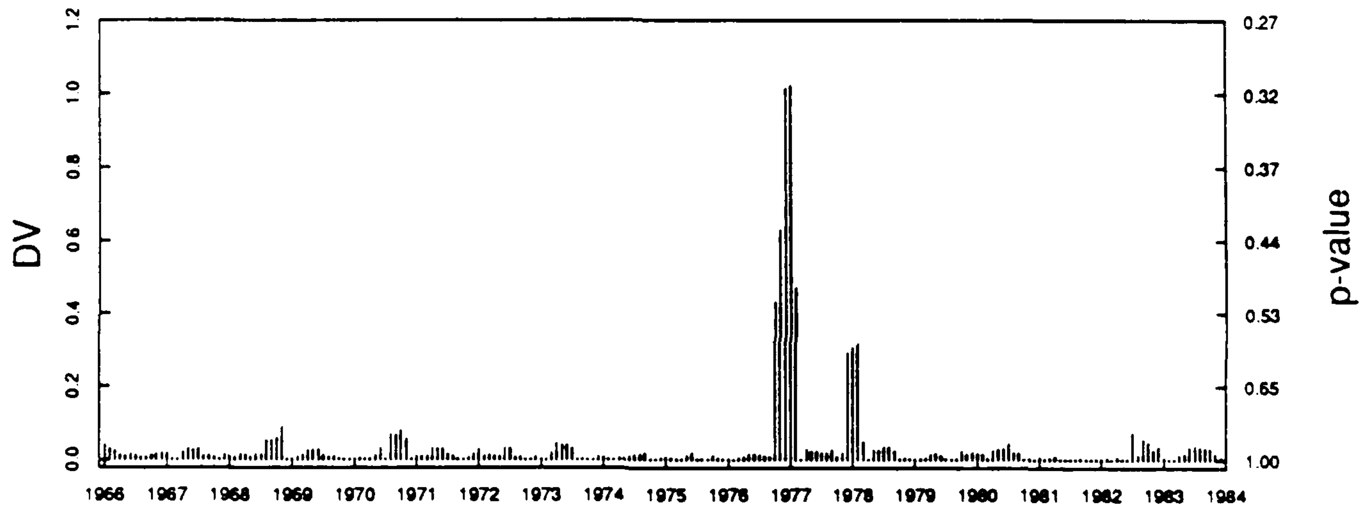
l) DV: Leave-2-Out after Two Rounds of Iterative Deletion

Example 7.1: Log of Exports to Latin America

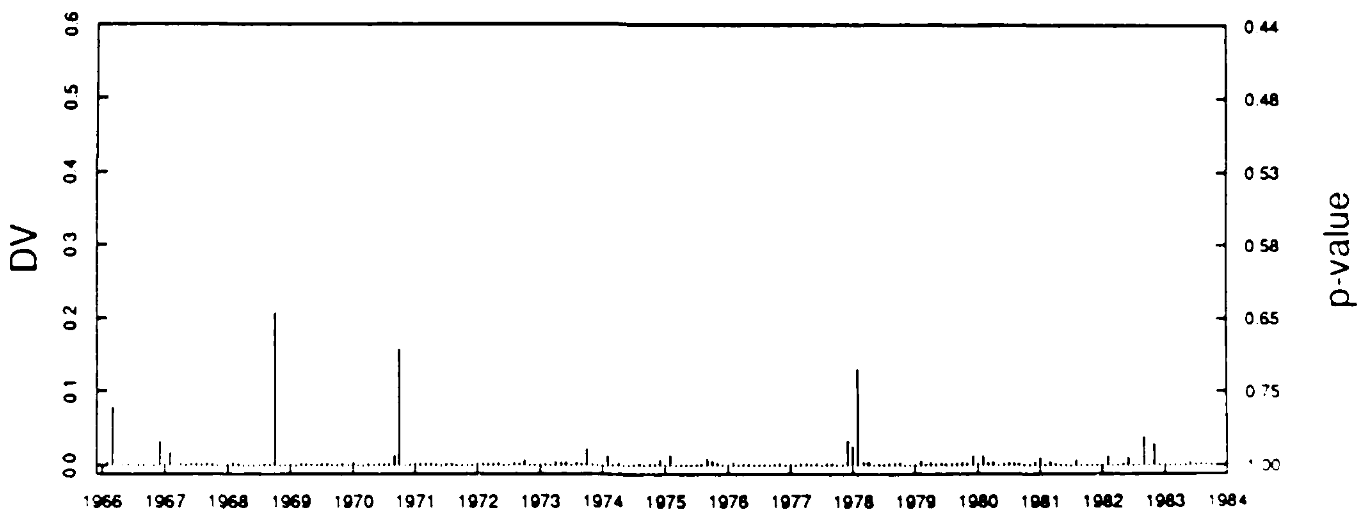
Figure 9



m) DV: Leave-3-Out after Two Rounds of Iterative Deletion



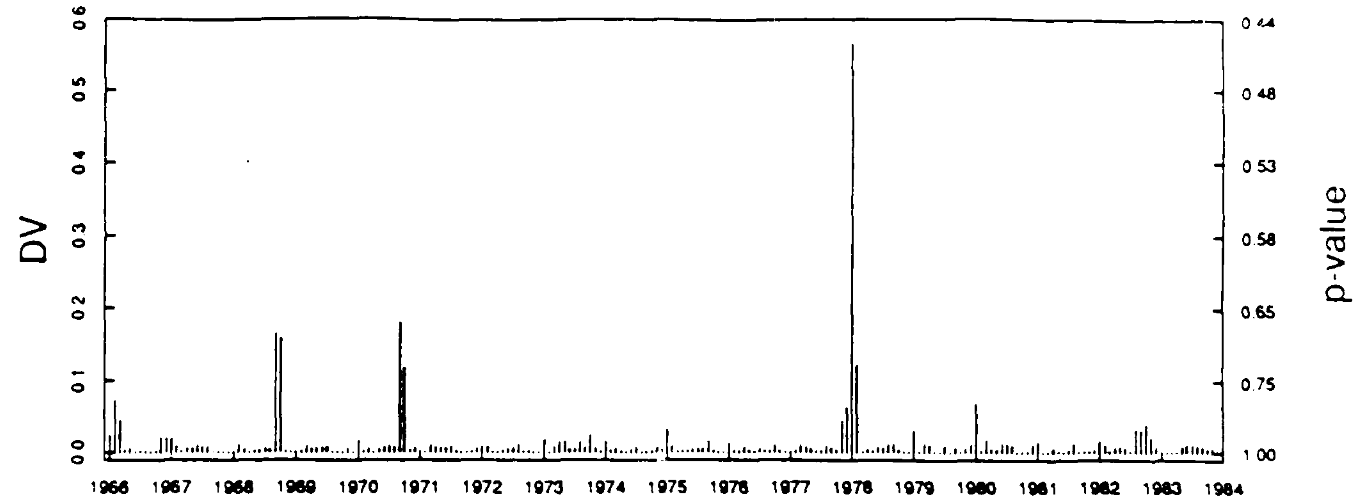
n) DV: Leave-4-Out after Two Rounds of Iterative Deletion



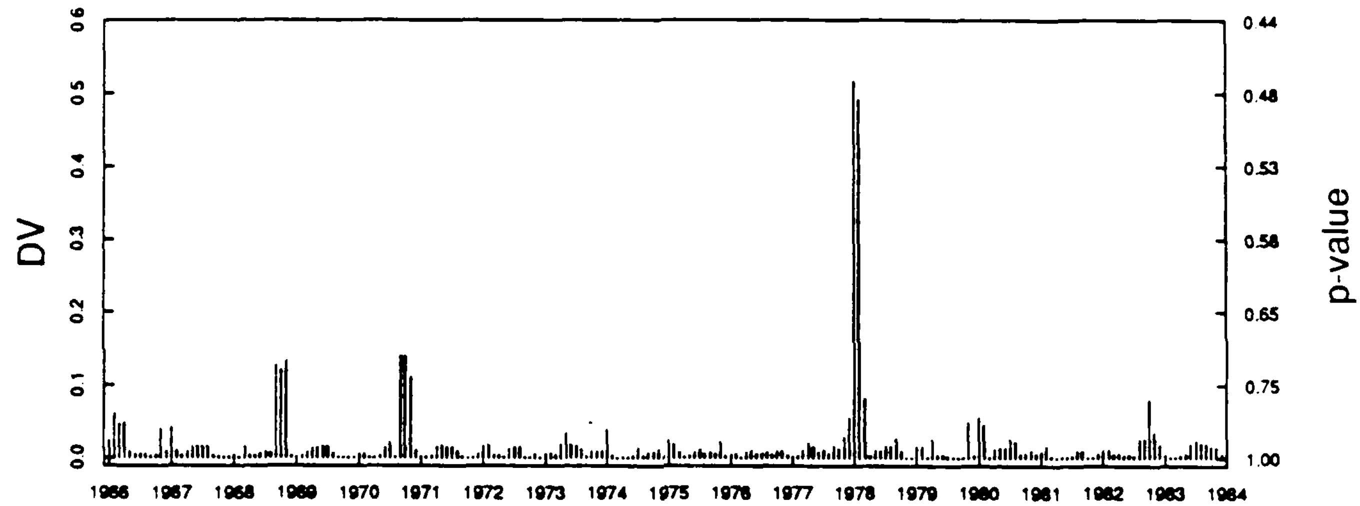
o) DV: Leave-1-Out after Three Rounds of Iterative Deletion

Example 7.1: Log of Exports to Latin America

Figure 9



p) DV: Leave-2-Out after Three Rounds of Iterative Deletion



q) DV: Leave-3-Out after Three Rounds of Iterative Deletion

Example 7.1: Log of Exports to Latin America

Figure 9

with the influential data points treated as missing provides a useful graphical display to be compared with the raw residual plot.

Example 7.2: Value of Unfilled Orders, Radio and TV (UNFTV)

Figure 10a displays the monthly value (in millions of dollars) of unfilled orders for radios and televisions (UNFTV) from 1958 to 1981. This series was previously studied by Martin, Samarov, and Vandaele (1983), who used a robust ACM Filter to fit an ARIMA$(0,1,1)\times(0,1,1)_{12}$. The series was also analyzed by Engle and Kraft (1983), who fit an ARCH model the data. Our initial fit is an ARIMA$(0,1,1)\times(0,1,1)_{12}$; the MLE's are given in Table 2 and the residuals from the MLE fit are given in Figure 10b.

Leave-k-out diagnostics for $k = 1,3,5$ are displayed in Figures 10c-10e. The diagnostics reveal a gross outlier at 9/78 (see Figure 10c), which apparently masks the influence of the other neighboring points: Figures 10d and 10e clearly indicate the presence of other influence observations at the end of the series. Following the iterative deletion strategy, we would remove the outlier at 9/78 and recompute the diagnostics. However, examination of Figure 10b shows that the end of series has many more large residuals than the rest of the series. It seems quite likely that a variance change may have occurred towards the end of the series. So instead of following the usual procedure, we adopt the flexible approach and look for the possibility of a variance shift.

Using the "bottom up" (and computationally expensive) approach described in Section 6, we perform leave-k-out diagnostics for $k = 16,32,64$. The diagnostics for $k = 64$ are displayed in Figure 10f, and dramatically support the conjecture of non-homogeneity of variance present in the series: the maximum value for DV is over 250 (note that this plot has a different scale from Figures 10c-10e). The diagnostics for $k = 16$ and $k = 32$ (not shown) display a similar pattern, although achieving a smaller maximum value. It is clear that the behavior of this series is fundamentally different towards the end of the data. Thus, following

step 3 of Section 7, the data is split into two series, and each part is analyzed separately. We chose 1/76 as the change point, based on the residuals plot and on computation of DV for a few judiciously selected subsets. Specifically, patches of increasing size were truncated from end the data, and the data was split (approximately) according to patch of maximum influence.

Checking the model order for the first part of the series again yielded an ARIMA$(0,1,1)\times(0,1,1)_{12}$ model. The MLE's for this model are given in Table 2. The residuals to this fit are given in Figure 10g. Note the reduction in the estimated innovations variance from 3123 to 1303. Leave-k-out diagnostics for $k = 2,4,8$, displayed in Figures 10h-10j, reveal several patches of influence. Two patches are espcially prominent: one during 1968 and another in 1972. The patch in 1972, which shows up only in the leave-8-out diagnostics, is associated with an obvious level shift spanning from 5/72-10/74 (see Figure 10a). The patch in 1968 corresponds to a large residual at 6/68, and has a less well defined structure. A local level shift is present during 11/67-5/68 or during 6/68-10/69, or both. In any case, the diagnostics help us identify problem areas in the data, and show that the large residual is associated with a patch of influential points rather than an isolated outlier.

A different model was selected for the latter section of the series: ARIMA$(0,1,1)\times(0,0,2)_6$ was fit, and the MLE's displayed in Table 2. The residuals are plotted in Figure 10k (with the abscissa rescaled appropriately); the circled points are eventually deleted from the series. The leave-1-out diagnostics, displayed in Figure 10*l*, reveal the isolated outlier at 9/78 that showed up prominently in the original set of diagnostics (see Figure 10c). However, a surprising feature pops up in the leave-k-out diagnostics for $k = 2,3,4$, given in Figures 10m-10o. A highly influential patch at 2/78-5/78 is discovered; leave-k-out for $k = 5$ (not shown) reveals no further significance. It is important to note that this patch does not correspond to any unusually large residuals (see Figure 10k).

Following the iterative deletion strategy, we remove the patch 2/78-5/78 and recompute the MLE's, which are given in Table 2. The estimated innovations variance drops from 8960 to 5340, and the estimated coefficients change dramatically, having a root on the unit circle. Recomputing the diagnostics (not shown) reveals the influential point at 9/78. Note that 9/78 was significant in the previous round, and the smearing was consistent with an isolated outlier (see Figures 11*l*-11o). Removal of 9/78 further reduces the estimated innovations variance, but has a relatively small effect on the estimated coefficients (see Table 2). No more influential points are uncovered with a second round of iterative deletion.

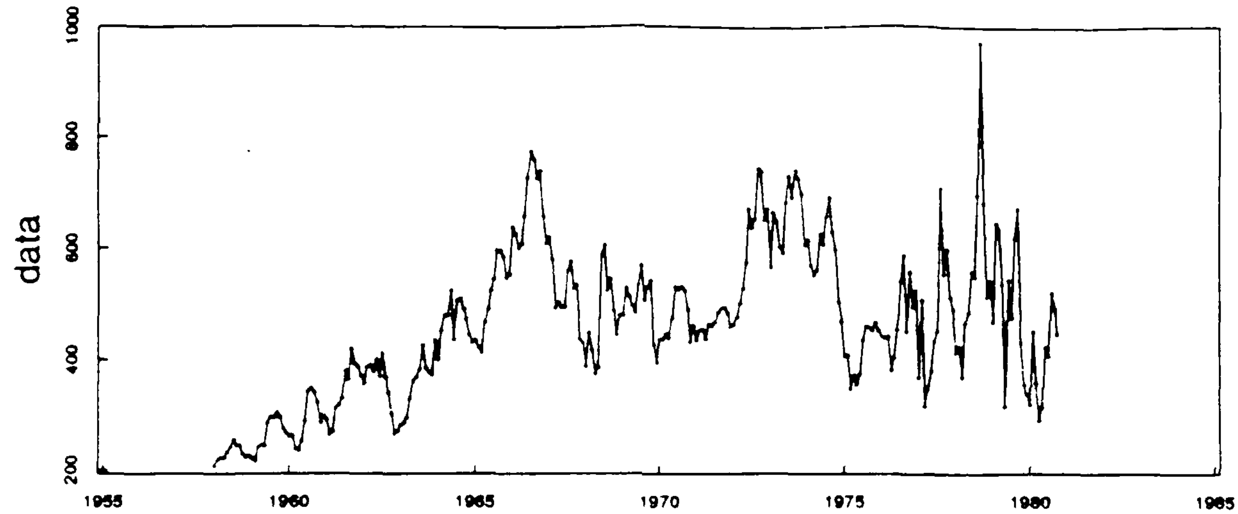
Table 2:
Parameters Fit to UNFTV Data

Time Period	Model	Step	$\hat{\theta}_1$	$\hat{\Theta}_1$	$\hat{\Theta}_2$	$\hat{\sigma}^2$	Points Deleted
1958-80	$(0,1,1)\times(0,1,1)_{12}$	—	.41	.75	—	3123	—
1958-75	$(0,1,1)\times(0,1,1)_{12}$	—	.18	.92	—	1303	—
	$(0,1,1)\times(0,0,2)_{6}$	0	.36	-.25	-.51	8960	—
1976-80	$(0,1,1)\times(0,0,2)_{6}$	1	.49	-.46	-1.00	5340	2/78-5/78
	$(0,1,1)\times(0,0,2)_{6}$	2	.36	-.46	-1.00	4173	9/78

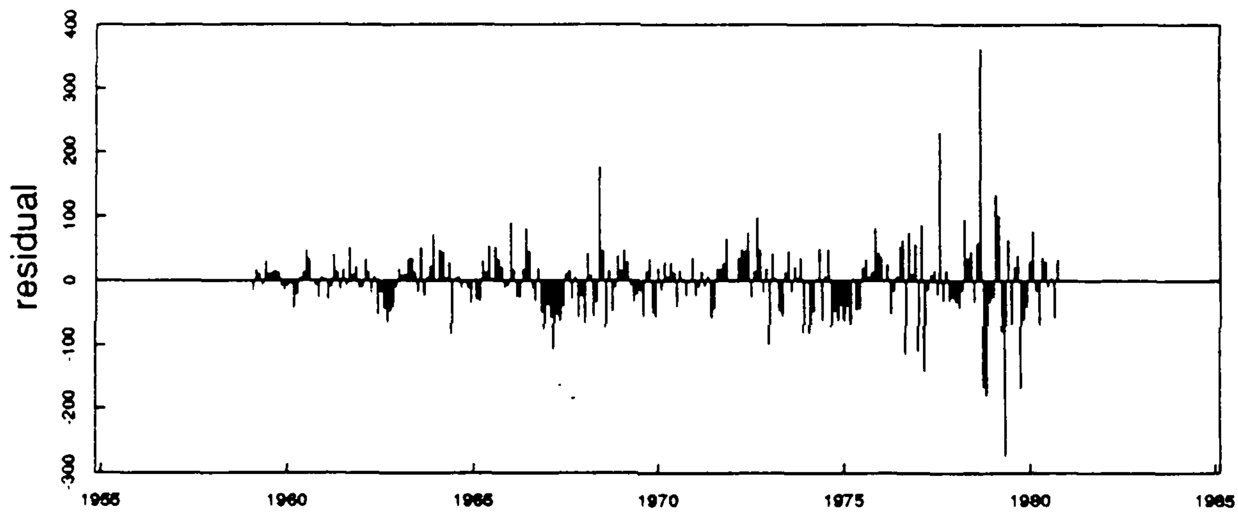
In summary, the UNFTV series clearly reveals the importance of leave-k-out diagnostics embedded in a good overall strategy! This approach effectively detects the major modeling difficulties present in the data. A single ARIMA model is inadequate to represent the entire series: the latter part of the data appears to behave according to a different model. Also, the first part of the series is subject to several local disturbances, which could be modeled as level shifts using intervention analysis. Finally, several very influential observations are present in the latter part of the data, affecting the estimated coefficients

dramatically. Obviously, any forecasts made will depend highly on how these points are treated.

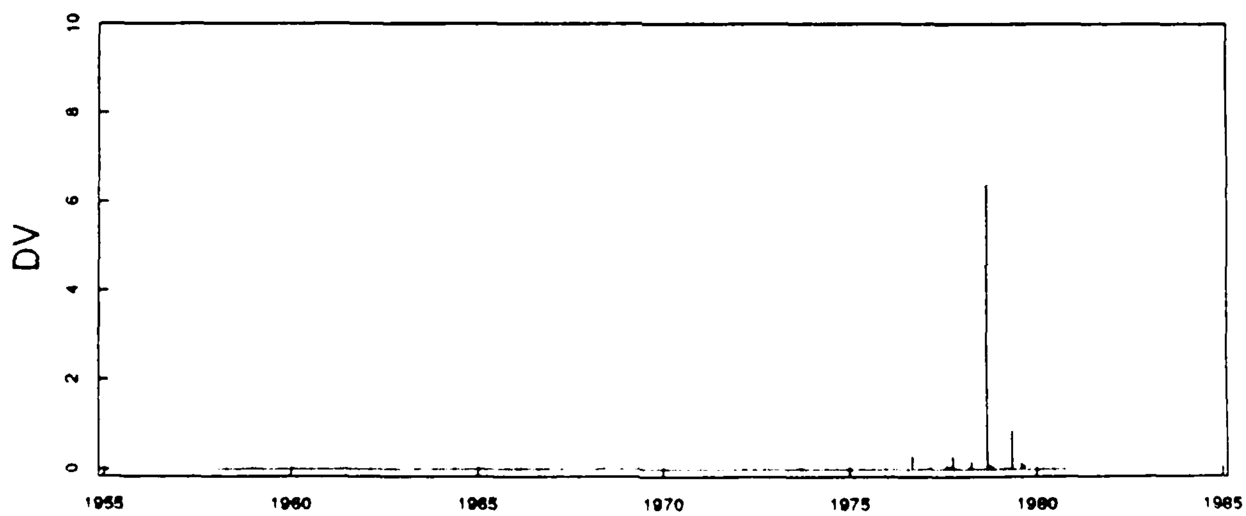
With regard to dealing with the variance shift in the last part of the series, the ARCH modeling approach of Engle and Kraft (1983) appears to be a viable alternative.



a) Plot of Data



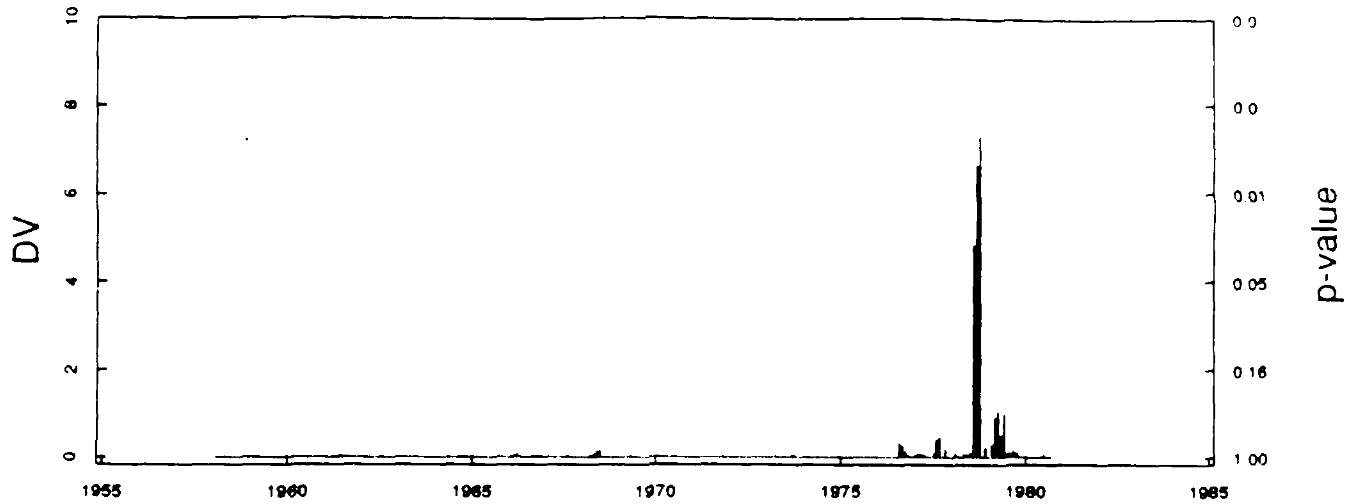
b) Plot of Residuals



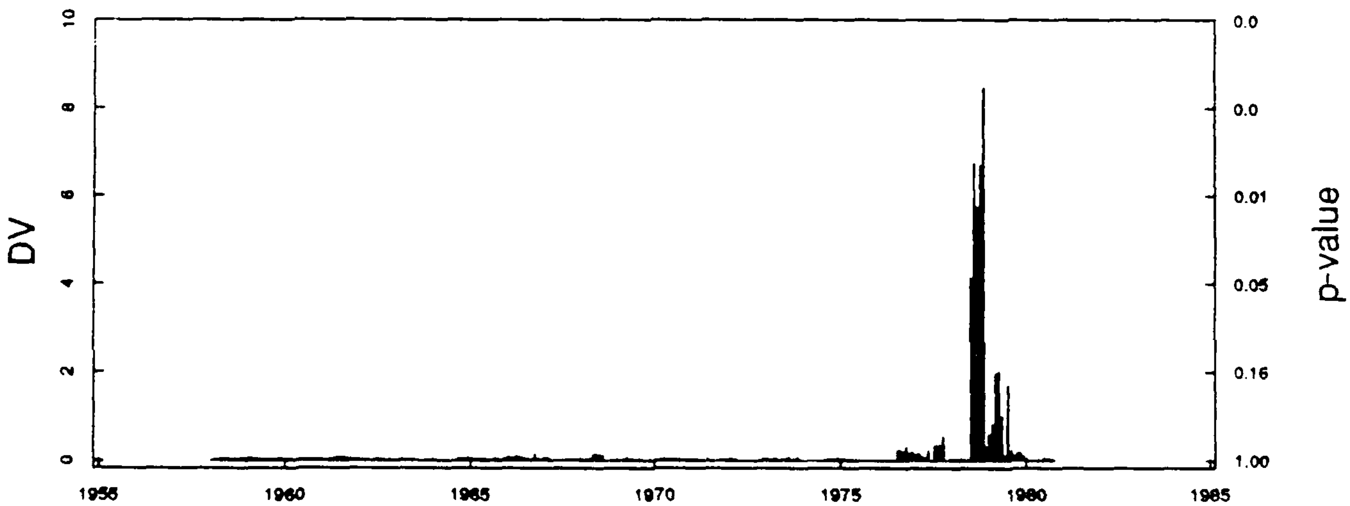
c) Scaled Leave-1-Out Diagnostics: Innovations Variance

Example 7.2: Unfilled TV Orders

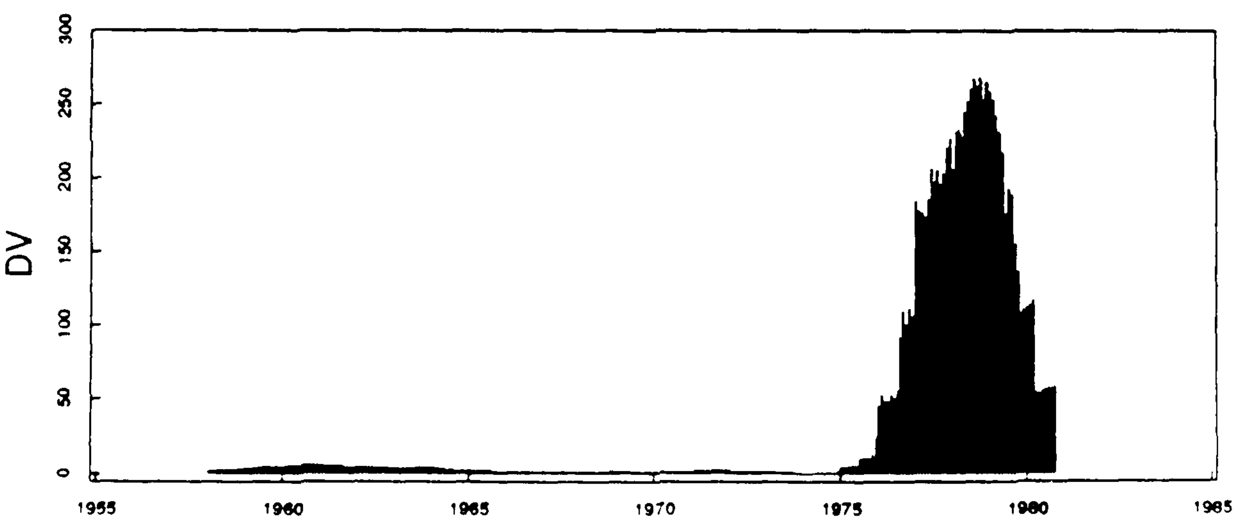
Figure 10



d) Scaled Leave-3-Out Diagnostics: Innovations Variance



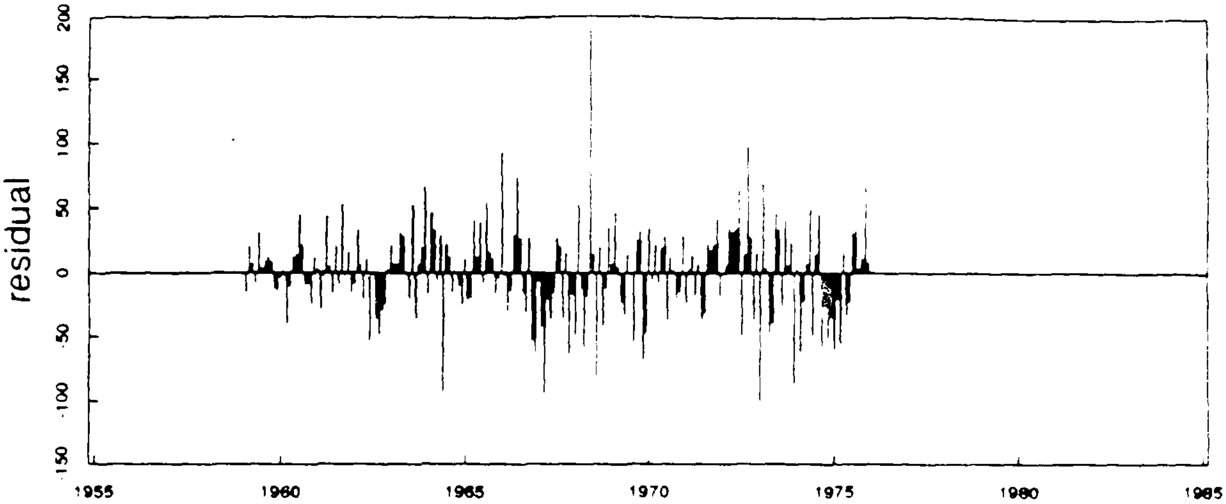
e) Scaled Leave-5-Out Diagnostics: Innovations Variance



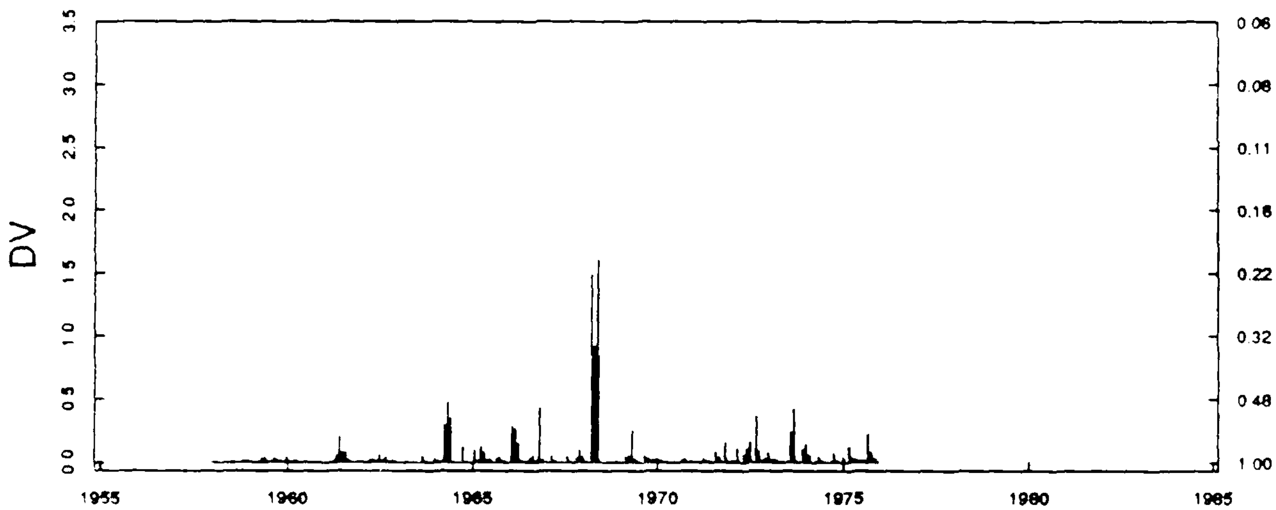
f) Scaled Leave-64-Out Diagnostics: Innovations Variance

Example 7.2: Unfilled TV Orders

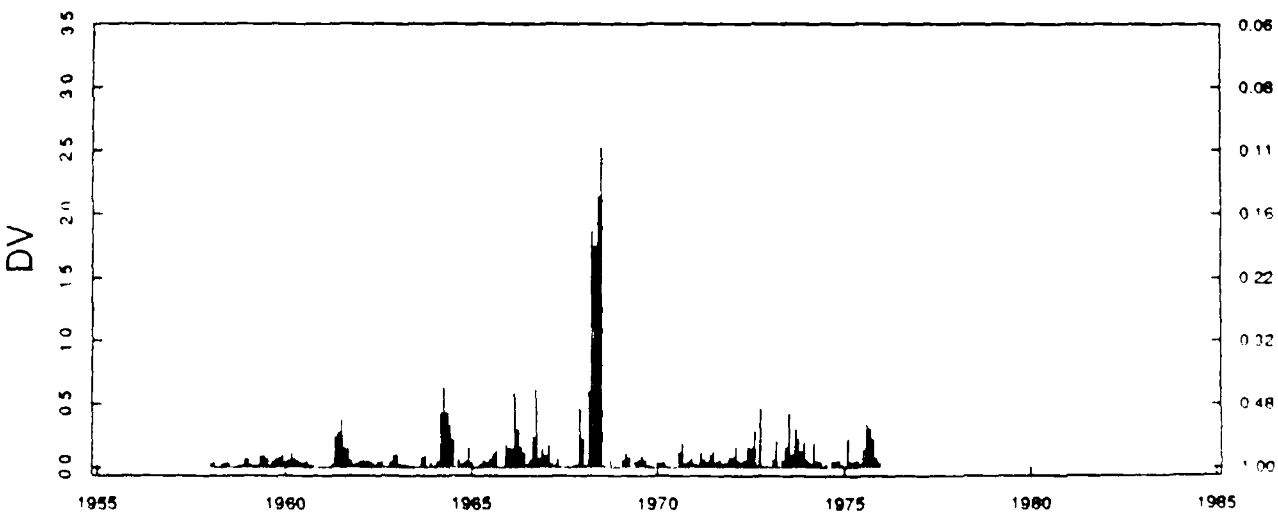
Figure 10



g) Plot of Residuals: 1958-1975



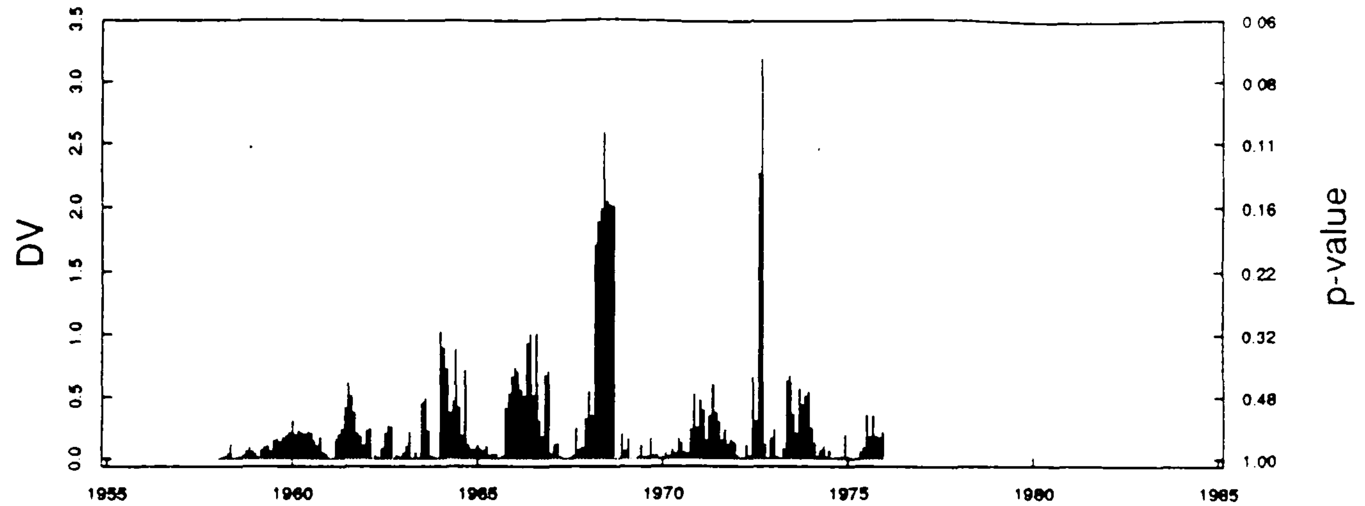
h) Scaled Leave-2-Out Diagnostics: Innovations Variance



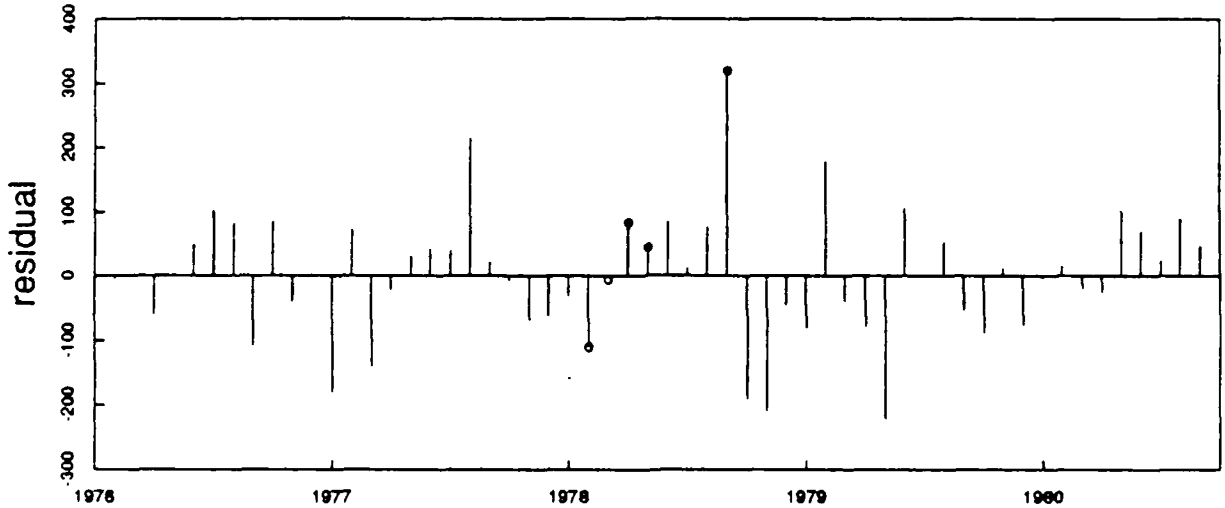
i) Scaled Leave-4-Out Diagnostics: Innovations Variance

Example 7.2: Unfilled TV Orders

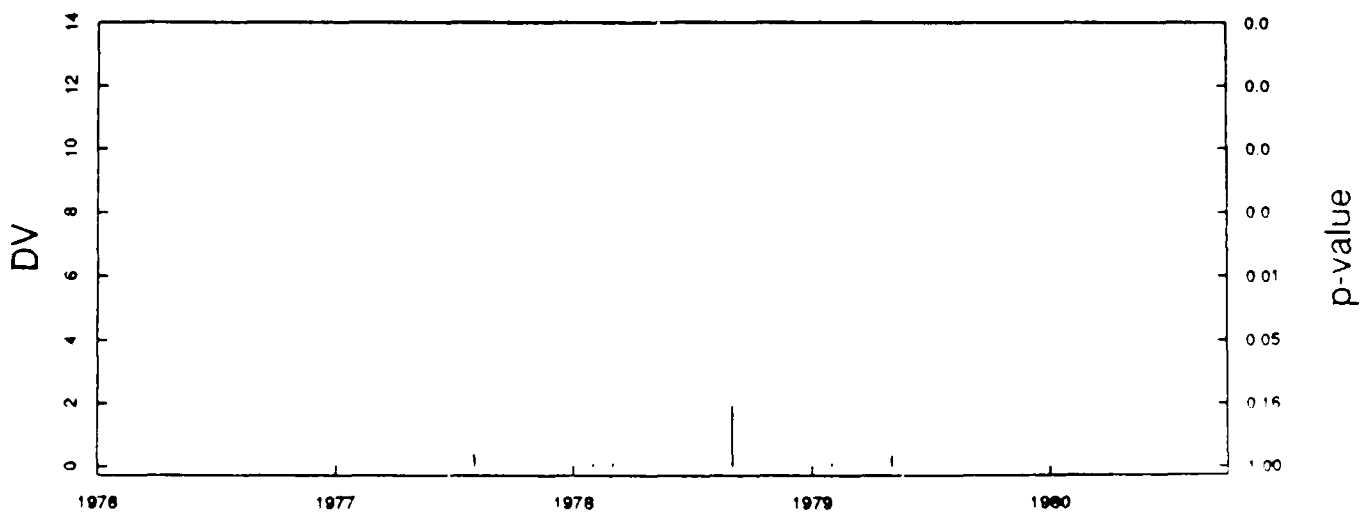
Figure 10



j) Scaled Leave-8-Out Diagnostics: Innovations Variance



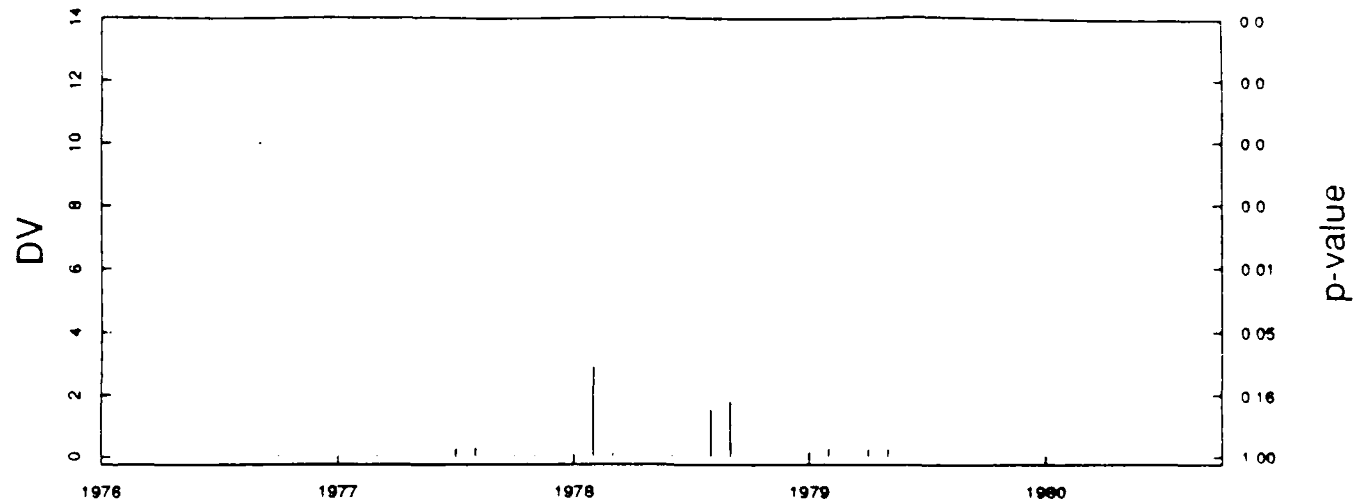
k) Plot of Residuals: 1976-1980



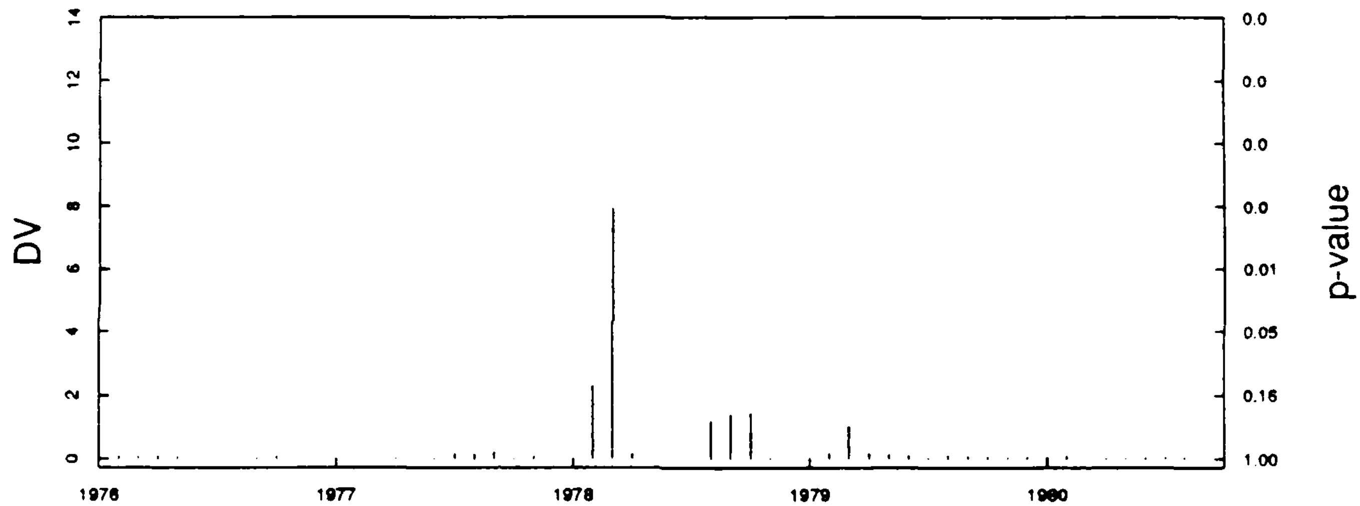
l) Scaled Leave-1-Out Diagnostics: Innovations Variance

Example 7.2: Unfilled TV Orders

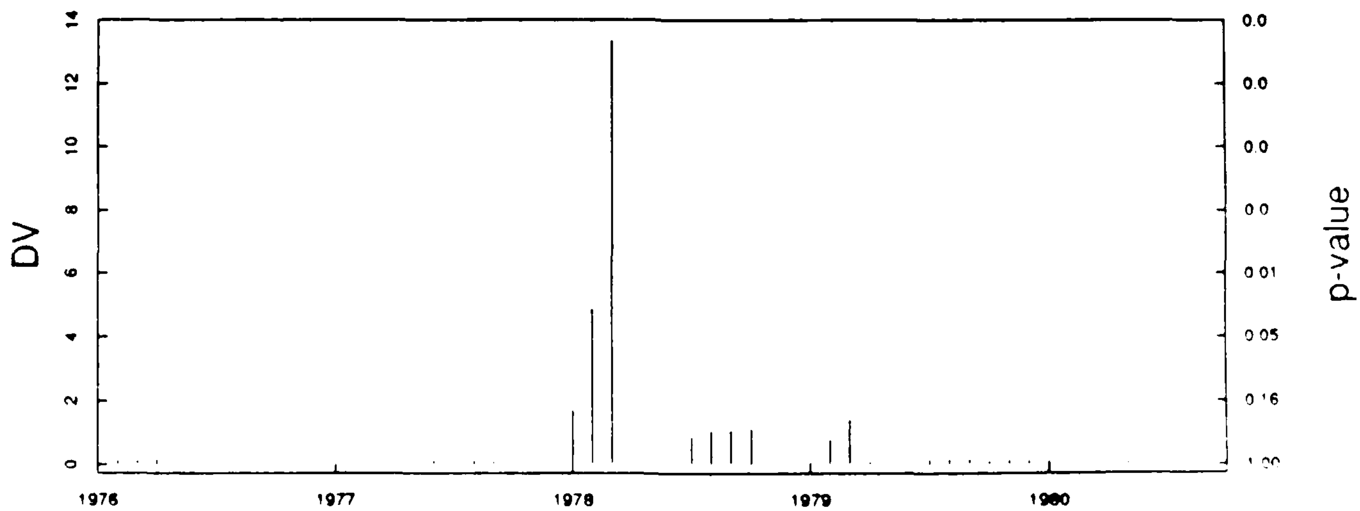
Figure 10



m) Scaled Leave-2-Out Diagnostics: Innovations Variance



n) Scaled Leave-3-Out Diagnostics: Innovations Variance



o) Scaled Leave-4-Out Diagnostics: Innovations Variance

Example 7.2 : Unfilled TV Orders

Figure 10

8. Discussion

8.1 Extensions: Other Measures of Influence

We have introduced two measures of empirical influence in this paper, based on the estimated coefficients and on the estimated innovations variance. While the former is more well established in the ordinary regression context, the latter was found to be more effective at identifying outliers and other types of influential observations. Other notions of influence may also be useful; several possibilities are discussed below.

Diagnostics Based on a Robust Scale Measure

One attractive extension, suggested by V. Yohai, consists of computing a robust scale measure of the residuals computed from the leave-k-out fit. To see why that would be useful, recall that an outlier will cause a large value of DV for two reasons: first, the outlier inflates the variance because of an associated large residual, and second, it inflates the variance by distorting the parameter estimates, and hence the fit. A robust measure of the scale of the residuals is resistant to outliers, and hence would reflect only the second feature. Thus, we obtain measure of how the fit alone is influenced by a patch of observations by computing the change in the robust scale for the leave-k-out residuals.

Diagnostics for Forcasting

An important area for application of subset deletion based diagnostics is in forecasting. One possibility is to measure the influence of a set of observations on the forecast distribution by computing the change in the location and spread of the forecast confidence intervals when subsets of points are deleted. It is expected that the most influential points for forecasting will occur near the end of the series. Interactive graphical displays of the change in the confidence intervals will be particularly useful for determining the extent to which such

influence exists.

Another possible measure of influence is to determine the change in forecast mean-square-error when points are deleted.

Diagnostics Based on Additive Noise Variance

We have demonstrated that the influence of an outlier in an ARIMA model is better measured by a diagnostic based on estimates of the innovations variance rather than a diagnostic based on the estimated coefficients. This raises the question whether or not a useful diagnostic might be based on estimates of hypothesized ***additive noise*** variance. The state space formulation of Section 2 easily generalizes to include additive noise in the measurement equation (see Jones, 1980). Let ζ_t be a sequence of normally distributed independent random variables with mean 0 and variance $h\,\sigma^2$. Then the measurement equation for an ARIMA process observed with additive noise is

$$x_t = \mathbf{z}'\mathbf{x}_t + \zeta_t \tag{9.1}$$

The prediction and update equations remain the same, except that the observation-prediction residual variance is given by

$$f_t = \mathbf{z}'\mathbf{P}_{t|t-1}\mathbf{z} + h\,. \tag{9.2}$$

The maximum likelihood estimate of h is obtained by inclusion of a parameter for h in the non-linear optimization of (2.10).

It is possible that inclusion of additive noise in the model will yield an even sharper diagnostic for outliers, and may be helpful in identifying certain model changes (e.g., level shifts). However, there is one caveat: Estimation of an additive noise variance $h\,\sigma^2$ when in fact $h\,\sigma^2 = 0$ can lead to seriously inflated variances for the other parameter estimates (see Section V of Martin, 1980).

8.2 Computational Considerations

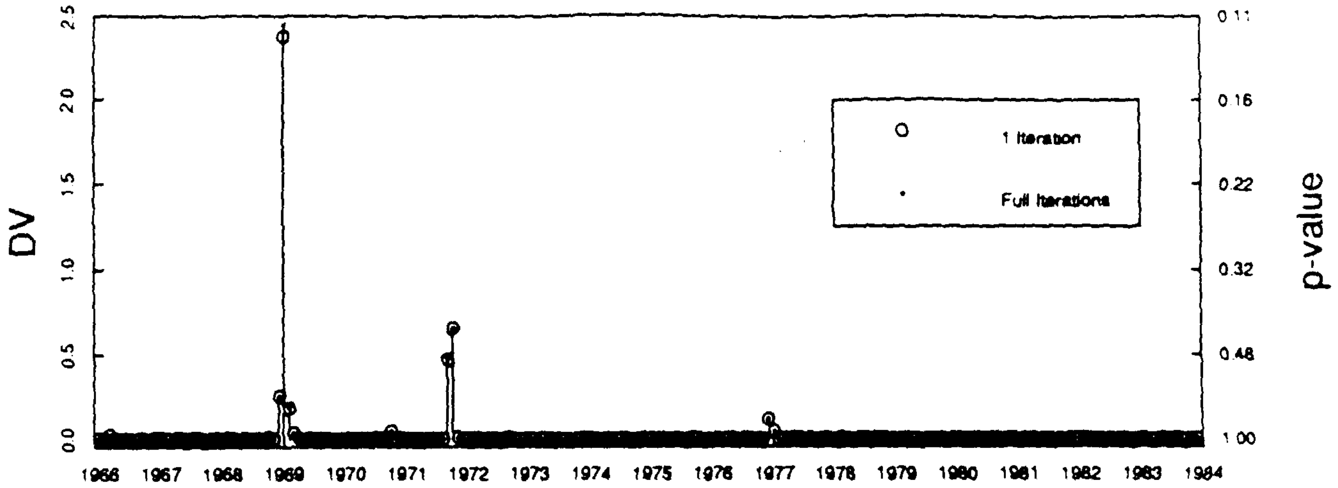
Software for computation of maximum likelihood estimates for ARIMA models with missing data is now widely available. Implementation of leave-k-out diagnostics involves a simple extension of existing programs or statistical packages. The complexity of leave-k-out diagnostics is $n \times j$ non-linear optimizations, where n is the sample size and j is the number of different k used for leave-k-out diagnostics. However, the computations are not as severe as it might appear, since the optimizations can be run with good starting values and an estimate of the hessian obtained from the fit with all the data.

The computational burden can be considerably reduced by iterating only one or two times in the optimization procedure (see Storer and Crowley, 1985). We have found this gives virtually identical results in many cases: See Figure 11 which reproduces the initial leave-one-out diagnostics for Example 7.1 (Figure 9d), and superimposes the diagnostics obtained when just one iteration is allowed.

Some run times in minutes for Examples 3.1 and 7.1 are given in Table 5. These computations were carried out on a 68020 based Masscomp MC5600 at the University of Washington. Restricting the optimizer to just one iteration reduces the computations by factors of roughly two and three in the two examples. Further improvements in speed are possible by computing analytical first and second derivatives.

The use of interactive graphics to mark subsets for deletion was mentioned as a tool for use when the iterative deletion procedure breaks down (see Section 6.2). Interactive subset deletions can also provide computational savings for those situations in which as much information can be gleaned from the data computing DV for a few select subsets as for the entire data set. Furthermore, implementation of such a procedure is relatively straightforward using current technology. We are developing implementions for **UNIX** † workstations

† UNIX is a trademark of AT&T Bell Laboratories.



a) Scaled Leave-1-Out Diagnostics: Full and 1 Iterations Allowed

Example 8.1 (continued): Log of Exports to Latin America
Diagnostics Computed Allowing Full Iterations and 1 Iteration

Figure 11

running the interactive statistical language and system S (Becker and Chanbers, 1984), and for **Symbolics** ‡ workstations.

Table 5

	Leave-k-out	No. of Iterations Allowed	CPU Time in minutes: seconds
Example 3.1 $n = 100$ *AR(1)*	$k = 1$	1 full	0:12 0:23
Example 7.1 $n = 216$ *ARIMA(0,1,2)*	$k = 1$	1 full	3:50 10:01
	$k = 5$	1 full	3:50 10:06

8.3 Scaling of Diagnostics

The diagnostic proposed in (3.5) for DC is not very satisfactory when the estimated coefficients are near the boundaries of nonstationarity or noninvertibility. This is because $\mathbf{I}(\alpha)$, the expected information matrix, becomes singular in this case, and inflates DC. The resulting diagnostic is no longer comparable to a χ^2 distribution, and is too heavily weighted by the coefficients responsible for the singularity. One solution to this problem is to scale by $\dot{\Psi}(\hat{\alpha})/n$, the observed information, which hopefully gives a better estimate of the covariance matrix for $\hat{\alpha}$ (Efron and Hinckley, 1978, prefer observed information in the case with independent observations). Computation of $\dot{\Psi}(\hat{\alpha})/n$ is more difficult, though it can be done explicitly. Kohn and Ansley (1986) give a state space formulation of the

‡ Symbolics is a trademark of Symbolics, Inc.

ARIMA model which includes a computation of the first and second derivatives.

We have not yet addressed the issue of "internal" versus "external" scaling of the diagnostics (see Cook and Weisberg, 1982). Internal scaling uses the same norm for all observations, while external scaling uses a different norm for each observation, with each norm based on the data excluding the observation. For simplicity, we have chosen to use an internal norm in (3.5) for DC. However, we could easily compute the externally scaled diagnostic

$$n\,(\hat{\alpha}-\hat{\alpha}_{k,t})'\hat{\mathbf{I}}_{k,t}(\alpha)(\hat{\alpha}-\hat{\alpha}_{k,t}) \tag{8.3}$$

where $\hat{\mathbf{I}}_{k,t}(\alpha)$ is the estimated information matrix for $\hat{\alpha}_{k,t}$. Two possible estimates are $\hat{\mathbf{I}}_{k,i}(\alpha)=\mathbf{I}(\hat{\alpha}_{k,t})$ or $\hat{\mathbf{I}}_{k,t}(\alpha)=\dot{\Psi}_{k,t}(\hat{\alpha}_{k,t})/(n-k)$. As in the case of ordinary regression analysis, internal scaling tends to obscure influential points, since an outlier will often inflate the variance of $\hat{\alpha}$, and hence will decrease $\mathbf{I}(\hat{\alpha})$. Thus, external scaling is preferable for DC, and should be used if possible.

For scaling the diagnostic DV based on the innovations (or observational) variance, it only makes sense to use an external norm. This is because an outlier virtually always increases the estimated variance, so that the ratio $\hat{\sigma}^2_{k,t}/\hat{\sigma}^2$ is usually less than one for outlying observations. Thus, if in (3.7) we scaled by the internal norm $\hat{\sigma}^4$, instead of the external norm $\hat{\sigma}^4_{k,t}$, then $DV(k,\cdot)$ is often not significant when evaluated for an outlier.

8.4 Diagnostics for the Beginning of the Series

In this paper, we have dealt with the beginning of the series by reversing the series so that the beginning becomes the end. The need for this arises because the Harvey and Pierse (1984) algorithm for missing data, which we have used, does not handle missing data at the beginning of the series. On the other hand, the method of Kohn and Ansley (1986) does allow missing values in the beginning of the series, and thus with their method, a complete set of (leave-k-out) diagnostics can be computed with one pass over the data. We hope to

implement the Kohn and Ansley method in the near future.

8.5 Diagnostics Versus Robust Filter and Smoother Cleaners

The results produced by the leave-k-out diagnostics are similar to the prediction residual diagnostics obtained from robust model fitting based on robust filter-cleaners or robust smoother-cleaners (see Martin, 1979, 1981; Martin and Thompson, 1982; Martin, Samarov and Van Daele, 1983). The robust model fitting diagnostics consist of looking for large values of the observation prediction residuals $e_t = y_t - \hat{y}_{t|t-1}$, where $\hat{y}_{t|t-1}$ denotes the one-step ahead robust prediction based on filter or smoother cleaned data.

There is in fact a close connection between the the prediction residuals produced by leave-k-out diagnostics and those produced by a hard rejection type filter-cleaner. Using the notation of Section 2, the hard rejection filter is defined by replacing the recursions for $\hat{\mathbf{x}}_t$ and $\mathbf{P}_t$ in (2.6) by

$$\hat{\mathbf{x}}_t = \begin{cases} \hat{\mathbf{x}}_{t|t-1} + f_t^{-1} \mathbf{P}_{t|t-1} \mathbf{z} \cdot v_t & \text{if } |e_t| \le c \\ \hat{\mathbf{x}}_{t|t-1} & \text{if } |e_t| > c \end{cases} \tag{8.4}$$

$$\mathbf{P}_t = \begin{cases} \mathbf{P}_{t|t-1} - f_t^{-1} \mathbf{P}_{t|t-1} \mathbf{z}\mathbf{z}' \mathbf{P}_{t|t-1} & \text{if } |e_t| \le c \\ \mathbf{P}_{t|t-1} & \text{if } |e_t| > c \end{cases} \tag{8.5}$$

where c is some threshold value (for hard rejection filtering, $c \approx 2.6$ works well; see Martin and Su, 1985). One way of looking at (8.4) is that a data value y_t corresponding to prediction residual larger in absolute value than c produces the same result as the Kalman filter with y_t treated as missing. Thus, if the iterative deletion procedure of Section 6.1 identifies the same points as those which are rejected by the filter of (8.4), then the prediction residuals of the two procedures will be identical if the model parameter values are the same. The latter will be approximately true at the completion of the overall strategy. Hence, the

diagnostics obtained from the leave-k-out procedure will closely match those resulting from a robust procedure based on the hard rejection filter.

8.6 Related Work

Other approaches to the problems of outliers and structural disturbances in time series have been explored in the literature. An approach proposed by Chang and Tiao (1982), Hillmer, Bell and Tiao (1983), and Tsay (1986) is based on iterative fitting of ARIMA models, utilizing Fox (1972) tests to decide whether an individual observation is an IO, AO, or not and outlier. The approach is easily extended to cover shifts in level and shifts in variance.

Another important direction for dealing with model changes of various types has been pursued by Harrison and Stevens (1976) and Smith and West (1983), who use a Bayesian approach. A mixture of normals is used to automatically adapt the model to outliers and other local structures. West (1986) and West, Harrison, and Migon (1986) propose a somewhat different method based on Bayes factors, in which a nominal model is compared to a "neutral" alternative.

APPENDIX A: Computation of Asymptotic Information Matrix

This section derives an analytical expression for the asymptotic information matrix of a stationary and invertible ARMA (p, q) process $\phi(B)x_t = \theta(B)\varepsilon_t$. The formulas are easily extended to nonstationary and seasonal models. Let $g_1, \cdots, g_p$ and $h_1, \cdots, h_q$ be the roots of the polynomials $\phi(B)$ and $\theta(B)$, so that

$$\begin{aligned}\phi(B) &= (1-g_1B)(1-g_2B)\cdots(1-g_pB)\\ \theta(B) &= (1-h_1B)(1-h_2B)\cdots(1-h_1B)\end{aligned} \qquad (A.1)$$

Assume that the roots are distinct: $g_i \neq g_j$ and $h_i \neq h_j$ for $i \neq j$.

Let c_i and d_i be the coefficients in the expansion of $\phi(B)^{-1}$ and $\theta(B)^{-1}$ respectively; i.e.,

$$\phi^{-1}(B) = \sum_{i=0}^{\infty} c_i B^i, \qquad \theta^{-1}(B) = \sum_{i=0}^{\infty} d_i B^i \qquad (A.2)$$

The asymptotic information matrix $\mathbf{I}(\alpha)$ is given by

$$\begin{aligned}
\mathbf{I}_{i,j} &= \sum_{k=0}^{\infty} c_k c_{k+j-i} && \text{if } 1 \le i \le j \le p\\
\mathbf{I}_{i,p+j} &= \sum_{k=0}^{\infty} d_k c_{k+j-i} && \text{if } 1 \le i \le p,\ 1 \le j \le q,\ i \le j\\
\mathbf{I}_{i,p+j} &= \sum_{k=0}^{\infty} c_k d_{k+j-1} && \text{if } 1 \le i \le p,\ 1 \le j \le q,\ j \le i\\
\mathbf{I}_{p+i,p+j} &= \sum_{k=0}^{\infty} d_k d_{k+j-1} && \text{if } 1 \le i \le j \le q
\end{aligned} \qquad (A.3)$$

The coefficients can be computed recursively from the relation

$1 = \phi(B)\phi(B)^{-1} = \theta(B)\theta(B)^{-1}$, or

$$\phi(B)c_t = 0 \quad \theta(B)d_t = 0 \quad t = 1, 2, \cdots \qquad (A.4)$$

Initial conditions for the recursions are $c_0 = 1$, $c_{-p+1} = \cdots = c_{-1} = 0$ and $d_0 - 1$, $d_{-p+1} = \cdots = d_{-1} = 0$. Hence, (A.3) provides and explicit expression for $\mathbf{I}(\alpha)$.

By expressing (A.3) in terms of the roots $\Phi(B)$ and $\Theta(B)$, the formulae can be reduced to a summation of a finite number of terms. For $t = 1, 2, \cdots$, a solution to (A.4) is given by (see Box Jenkins [1976])

$$\begin{aligned} c_t &= k_1 g_1^t + \cdots + k_p g_p^t \\ d_t &= l_1 h_1^t + \cdots + l_q h_q^t \end{aligned} \tag{A.5}$$

where $k_1, \cdots, k_p$ and $l_1, \cdots, l_q$ are (possibly complex valued) constants. Since c_t and d_t can be evaluated recursively, (A.5) defines a system of linear equations which can be solved for $k_1, \cdots, k_p$ and $l_1, \cdots, l_q$. Substituting (A.5) into (A.3), Fubini's theorem yields

$$\begin{aligned} \mathbf{I}_{i,j} &= \sum_{t=0}^{\infty} \Big(\sum_{m=1}^{p} k_m g_m^{t+j-i} \Big) \Big(\sum_{n=1}^{p} k_n g_n^t \Big) \\ &= \sum_{m=1}^{p} \sum_{n=1}^{p} \Big(k_m k_n g_m^{j-i} \sum_{t=0}^{\infty} (g_m g_n)^t \Big) \\ &= \sum_{m=1}^{p} \sum_{n=1}^{p} \Big(k_m k_n g_m^{j-i} (1 - g_m g_n)^{-1} \Big) \quad \text{if } i \le j \end{aligned} \tag{A.6a}$$

Similarly,

$$\mathbf{I}_{i,p+j} = \sum_{m=1}^{p} \sum_{n=1}^{p} \Big(k_m l_n g_m^{j-i} (1 - g_m h_n)^{-1} \Big) \quad \text{if } i \le j \tag{A.6b}$$

$$\mathbf{I}_{i,p+j} = \sum_{m=1}^{p} \sum_{n=1}^{p} \Big(k_m l_n h_m^{j-i} (1 - g_m h_n)^{-1} \Big) \quad \text{if } j \le i \tag{A.6c}$$

$$\mathbf{I}_{p+i,p+j} = \sum_{m=1}^{p} \sum_{n=1}^{p} \Big(k_m l_n h_m^{j-i} (1 - h_m h_n)^{-1} \Big) \quad \text{if } i \le j \tag{A.6d}$$

APPENDIX B: Computation of EDC

The expected asymptotic diagnostics for coefficients, EDC, are computed for the leave-one-out diagnostic in the AR(1) model. First, we compute $\Delta(t_0; \phi)$, the difference between score function for the entire data and the score function for the data with t_0 treated as missing for a perfectly observed Gaussian process. Then, for the AO and IO contamination models studied in Section 4, (4.6) is evaluated by breaking up the computations into two parts, one corresponding to the outlier free process, and the other corresponding to the contamination.

Let x_t be a perfectly observed Gaussian process. For the AR(1) process, the score function for ϕ is

$$\Psi(\phi) = -\frac{1}{2}\sum_{t=1}^{N} \dot{f}_t / f_t - \frac{1}{2\sigma^2}\sum_{t=1}^{N} [2e_t \dot{e}_t / f_t - e_t^2 \dot{f}_t / f_t^2] \tag{B.1}$$

where e_t is the prediction residual, $\sigma^2 f_t$ is the variance of the prediction residual, $\dot{e}_t = \dfrac{\partial e_t}{\partial \phi}$, and $\dot{f}_t = \dfrac{\partial f_t}{\partial \phi}$ With no missing data, we have $e_t = x_t - \phi x_{t-1}$ and $f_t = 1$ for $t > 1$. If x_{t_0} is missing, the score function is similar to (B.1), except that we drop the t_0 term in the summations and adjust the $t_0 + 1$ prediction and residual. Define $e^*_{t_0} = 0$ and $f^*_{t_0} = 1$, and denote the prediction residuals and variances for when x_{t_0} is missing by e_t^* and $\sigma^2 f_t^*$. If $t_0 > 1$, then

$$e^*_{t_0+1} = x_{t_0+1} - \phi^2 x_{t_0-1} \qquad f^*_{t_0+1} = (1+\phi^2) \tag{B.2}$$

and $e_t^* = e_t$, $f_t^* = f_t$ for $t \neq t_0, t_0+1$. The score function $\Psi^{(t_0)}(\phi)$, with x_{t_0} missing, is given by (B.1) with e_t^* and f_t^* substituted for e_t and f_t.

Hence for $t_0 > 1$

$$\Delta_{t_0}(t_0; \phi) = \Psi_n^{(t_0)}(\phi) - \Psi_n(\phi)$$

$$= -\frac{1}{2}\frac{\dot{f}^*_{t_0+1}}{f^*_{t_0+1}} - \frac{1}{2\sigma^2}[2e^*_{t_0+1}\dot{e}^*_{t_0+1}/f^*_{t_0+1} - e^{*^2}_{t_0+1}\dot{f}^*_{t_0+1}/f^{*^2}_{t_0+1}]$$

$$+ \frac{1}{\sigma^2}[e_{t_0}\dot{e}_{t_0} + e_{t_0+1}\dot{e}_{t_0+1}]$$

$$= -\frac{\phi}{1+\phi^2} - \frac{1}{2\sigma^2}\left[\frac{2(x_{t_0}-\phi^2 x_{t_0-1})(-2\phi x_{t_0-1})}{1+\phi^2} - \frac{(x_{t_0+1}-\phi^2 x_{t_0-1})^2(2\phi)}{(1+\phi^2)^2}\right]$$

$$-\frac{1}{\sigma^2}\left[(x_{t_0}-\phi x_{t_0-1})x_{t_0-1} + (x_{t_0+1}-\phi x_{t_0})x_{t_0}\right]$$

$$= -\frac{\phi}{1+\phi^2} - \frac{1}{\sigma^2}\left[x_{t_0}(x_{t_0-1}+x_{t_0+1}) + x_{t_0-1}x_{t_0+1}\left[\frac{2\phi^3-2\phi(1+\phi^2)}{(1+\phi^2)^2}\right]\right.$$

$$\left. + x_{t_0-1}^2\left[\frac{2\phi^3(1+\phi^2)-\phi^5-(1+\phi^2)^2\phi}{(1+\phi^2)^2}\right] - \phi x_{t_0}^2 - x_{t_0+1}^2\left[\frac{\phi}{(1+\phi^2)^2}\right]\right]$$

$$= -\frac{\phi}{1+\phi^2} - \frac{1}{\sigma^2}\left[-\phi x_{t_0}^2 + x_{t_0}(x_{t_0-1}+x_{t_0+1}) - \frac{\phi}{(1+\phi^2)^2}(x_{t_0-1}+x_{t_0+1})^2\right] \quad \text{(B.3)}$$

which yields (4.10).

We are now set to compute *EDC* for the AR(1) model. Let y_t be contaminated according to an AO or IO outlier model. It is convenient to break up the computation of *EDC* into two parts, that corresponding to x_t and that corresponding to the contamination. Define $C_\phi \equiv E\,\Delta^2(t_0;\phi)$ where $\Delta(t_0;\phi)$ is the difference in the score functions for x_t, and is given by (4.10). Denote, for now, the difference in the score functions for y_t by $\Delta_{y_t}(t;\phi)$, and let $\delta_{y_t}(t;\phi) \equiv \Delta_{y_t}(t;\phi) - \Delta(t;\phi)$. Then

$$E\,\Delta_{y_t}(t;\phi)^2 = C_\phi + E\,\delta_{y_t}(t;\phi)^2 \quad \text{(B.4)}$$

since the cross product terms vanish. From (B.4), we can compute *EDC* by equation (4.6).

The tedious part of (B.4) is computing C_ϕ, which is shown below to be equal to $\frac{2}{(1+\phi^2)^2}$. Calculating $E\,\delta_{y_t}(t;\phi)^2$ is easier, but must be done case by case. These are computed below for several outlier configurations.

AO Models

Consider first $\mathbf{y}_t$ which obeys the model given in (4.11): a single AO type outlier of size ζ at t_0. Paralleling the notation of Section 4, denote $\delta_{\mathbf{y}_t}(t_0;\phi)$ by $\delta^{AO}_{(\zeta;t_0)}(t_0;\phi)$. We obtain $\delta^{AO}_{(\zeta;t_0)}(t_0;\phi)$ by subtracting (4.10) from (4.12), which leads to

$$E\,(\delta^{AO}_{(\zeta;t_0)}(t_0;\phi)^2) = \frac{1}{\sigma^4}E\,[\zeta(x_{t_0-1}+x_{t_0+1}-2\phi x_{t_0})-\phi\zeta^2]^2$$

$$= \left[\frac{\zeta}{\sigma}\right]^4\phi^2+\left[\frac{\zeta}{\sigma}\right]^2\frac{1}{\sigma^2}E\,[\,4\phi^2x_{t_0}^2-4\phi x_{t_0}(x_{t_0-1}+x_{t_0+1})+(x_{t_0-1}+x_{t_0+1})^2\,]$$

$$= \left[\frac{\zeta}{\sigma}\right]^4\phi^2+\left[\frac{\zeta}{\sigma}\right]^2\frac{1}{1-\phi^2}(4\phi^2-8\phi^2+2(1+\phi^2))$$

$$= \left[\frac{\zeta}{\sigma}\right]^4\phi^2+\left[\frac{\zeta}{\sigma}\right]^2 2 \tag{B.5}$$

Upon adding C_ϕ and scaling by $(1-\phi^2)$, we get $EDC^{AO}_{(\zeta;t_0)}(t_0;\phi)$, which is displayed in (4.13).

Under the same outlier model, except leaving t_0+1 (or t_0-1) out, we get

$$E\,(\delta^{AO}_{(\zeta;t_0)}(t_0+1;\phi)^2) = \frac{1}{\sigma^4}E\left[-\zeta^2\frac{\phi}{(1+\phi^2)^2}+\zeta\,(x_{t_0}-\frac{2\phi}{(1+\phi^2)^2}(x_{t_0-1}+x_{t_0+1}))\right]^2$$

$$= \left[\frac{\zeta}{\sigma}\right]^4\frac{\phi^2}{(1+\phi^2)^4}+\left[\frac{\zeta}{\sigma}\right]^2\frac{1}{1-\phi^2}\left[1-\frac{4\phi}{(1+\phi^2)^2}(2\phi)+\frac{4\phi^2}{(1+\phi^2)^4}(2(1+\phi^2))\right]$$

$$= \left[\frac{\zeta}{\sigma}\right]^4\frac{\phi^2}{(1+\phi^2)^4}+\left[\frac{\zeta}{\sigma}\right]^2\frac{1}{1-\phi^2}\left[1-\frac{8\phi^4}{(1+\phi^2)^3}\right] \tag{B.6}$$

which leads to (4.14).

For AO type outliers of size ζ at both t_0-1 and t_0+1,

$$E\,\delta^{AO}_{(\zeta;t_0-1,t_0+1)}(t_0;\phi)^2 = \frac{1}{\sigma^4}E\left[-(2\zeta)^2\frac{\phi}{(1+\phi^2)^2}+(2\zeta)\left[x_{t_0}-2\frac{\phi}{(1+\phi^2)^2}(x_{t_0-1}+x_{t_0+1})\right]\right]^2$$

$$= E\,\delta^{AO}_{(2\zeta;t_0)}(t_0-1;\phi)^2 \qquad \text{(B.7)}$$

which is given by (B.6) above. Hence $\delta^{AO}_{(\zeta; t_0-1, t_0+1)}(t_0; \phi) = EDC^{AO}_{(2\zeta; t_0)}(t_0 - 1)$.

IO Models:

Now suppose y_t is observed with an IO type outlier of size ζ at t_0 and thus behaves according to (4.18). We obtain an expression for $\delta^{IO}_{(\zeta, t_0)}(t_0)$ by subtracting (4.10) from (4.18), so that

$$E\,\delta^{IO}_{(\zeta, t_0)}(t_0; \phi)^2 = \frac{1}{\sigma^4} E \left[\zeta^2 \left[-\phi + \phi - \frac{\phi^3}{(1+\phi^2)^2} \right] \right.$$

$$\left. + \zeta \left[x_{t_0}(-2\phi + \phi) + (x_{t_0-1} + x_{t_0+1}) \left[1 - \frac{2\phi^2}{(1+\phi^2)^2} \right] \right] \right]^2$$

$$= \left[\frac{\zeta}{\sigma}\right]^4 \frac{\phi^6}{(1+\phi^2)^4} + \left[\frac{\zeta}{\sigma}\right]^2 \frac{1}{1-\phi^2} \left[\phi^2 - 2\phi \left[1 - \frac{2\phi^2}{(1+\phi^2)^2} \right] (2\phi) \right.$$

$$\left. + \left[1 - \frac{2\phi^2}{(1+\phi^2)^2} \right]^2 \left[2(1+\phi^2) \right] \right]$$

$$= \left[\frac{\zeta}{\sigma}\right]^4 \frac{\phi^6}{(1+\phi^2)^4} + \left[\frac{\zeta}{\sigma}\right]^2 \frac{1}{1-\phi^2} \left[\phi^2 - 4\phi^2 + \frac{8\phi^4}{(1+\phi^2)^2} + 2(1+\phi^2) \right.$$

$$\left. - \frac{8\phi^2}{(1+\phi^2)} + \frac{8\phi^4}{(1+\phi^2)^3} \right]$$

$$= \left[\frac{\zeta}{\sigma}\right]^4 \frac{\phi^6}{(1+\phi^2)^4} + \left[\frac{\zeta}{\sigma}\right]^2 \frac{1}{1-\phi^2} \left[2 - \phi^2 \left[1 - \frac{8\phi^2(1+\phi^2) - 8(1+\phi^2)^2 + 8\phi^2}{(1+\phi^2)^3} \right] \right]$$

$$= \left[\frac{\zeta}{\sigma}\right]^4 \frac{\phi^6}{(1+\phi^2)^4} + \left[\frac{\zeta}{\sigma}\right]^2 \frac{1}{1-\phi^2} \left[2 - \phi^2 \left[1 + \frac{8}{(1+\phi^2)^3} \right] \right] \qquad \text{(B.8)}$$

From this follows the expression for $EDC^{IO}_{(\zeta; t_0)}(t_0)$, displayed in (4.20).

Evaluating the diagnostic at $t_0 + i$, we obtain $\delta^{IO}_{(\zeta, t_0)}(t_0 + i)$ from (4.10) by replacing x_{t_0-1} with $x_{t_0-1} + \phi^{i+1}\zeta$, x_{t_0} with $x_{t_0} + \phi^i \zeta$ and x_{t_0+1} with $x_{t_0+1} + \phi^{i-1}\zeta$ in (4.10). This yields

$$E\,\delta^{IO}_{(\zeta, t_0)}(t_0 + i)^2 = \frac{1}{\sigma^4} E \left[\zeta^2 \left[-\phi^{2i+1} + \phi^{2i+1} + \phi^{2i-1} - \frac{\phi}{(1+\phi^2)^2} (\phi^{i+1} + \phi \right. \right.$$

$$+\zeta\left[x_{t_0}(-2\phi^{i+1}+\phi^{i+1}+\phi^{i-1})+(x_{t_0-1}+x_{t_0+1})\left[\phi^i-\frac{2\phi}{(1+\phi^2)^2}(\phi^{i+1}+\phi^{i-1})\right]\right]^2$$

$$=\left[\frac{\zeta}{\sigma}\right]^4\phi^{2i-2}\left[\phi-\frac{\phi}{(1+\phi^2)^2}(\phi^2+1)^2\right]$$

$$+\left[\frac{\zeta}{\sigma}\right]^2\frac{1}{1-\phi^2}\phi^{2i-2}\left[(1-\phi^2)^2+2(1-\phi^2)\left[\phi-\frac{2\phi}{(1+\phi^2)}\right](2\phi)\right.$$

$$\left.+\left[\phi-\frac{2\phi}{(1+\phi^2)}\right]^2(2(1+\phi^2))\right]$$

$$=\left[\frac{\zeta}{\sigma}\right]^2\frac{1}{1-\phi^2}\phi^{2i-2}\left[(1-\phi^2)^2-\frac{4\phi^2(1-\phi^2)^2}{(1+\phi^2)}+\frac{2\phi^2(1-\phi^2)^2}{(1+\phi^2)}\right]$$

$$=\left[\frac{\zeta}{\sigma}\right]^2\phi^{2i-2}\frac{(1-\phi^2)}{(1+\phi^2)}(1+\phi^2-4\phi^2+2\phi^2)$$

$$=\left[\frac{\zeta}{\sigma}\right]^2\phi^{2i-2}\frac{(1-\phi^2)^2}{(1+\phi^2)} \tag{B.9}$$

This leads to (4.21).

Computation of C_ϕ:

It remains only to compute $C_\phi \equiv E\,\Delta_{t_0}(\phi;\mathbf{x}_t)^2$. Since the vector (x_{t-1}, x_t, x_{t+1}) has a multivariate normal distribution, we can easily compute the following useful expectations:

$$E\,[x_t^3(x_{t-1}+x_{t+1})] = 6\phi\frac{\sigma^4}{(1-\phi^2)^2} \tag{B.10}$$

$$E\,[x_t^2(x_{t-1}+x_{t+1})^2] = 2E\,[x_t^2x_{t-1}^{\zeta}+x_{t-1}x_t^2x_{t+1}]$$

$$= 2(1+5\phi^2)\frac{\sigma^4}{(1-\phi^2)^2}$$

$$E\,[x_t(x_{t-1}+x_{t+;})^3] = 2E\,[x_tx_{t-1}^3+3x_{t-1}^2x_tx_{t+1}]$$

$$= 2[3\phi+3\phi(1+2\phi^2)]\frac{\sigma^4}{(1-\phi^2)^2}$$

$$= 12\phi(1+\phi^2)\frac{\sigma^4}{(1-\phi^2)^2}$$

$$E[(x_{t-1}+x_{t+1})^4] = E[2(x_{t-1}^4+4x_{t-1}^3x_{t+1})+6x_{t-1}^2x_{t+1}^2]$$

$$= [2(3+12\phi^2)+6(1+2\phi^4)]\frac{\sigma^4}{(1-\phi^2)^2}$$

$$= 12(1+\phi^2)^2\frac{\sigma^4}{(1-\phi^2)^2}$$

We are now set to evaluate the expectation of the second term (squared) in (4.10).

$$E\left[-\phi x_t^2 + x_t(x_{t-1}+x_{t+1}) - \frac{\phi}{(1+\phi^2)^2}(x_{t-1}+x_{t+1})^2\right]^2$$

$$= E\left[\phi^2 x_t^4 - 2\phi x_t^3(x_{t-1}+x_{t+1}) + \left[\frac{2\phi^2}{(1+\phi^2)^2}+1\right]x_t^2(x_{t-1}+x_{t+1})^2\right.$$

$$\left. -2\frac{\phi}{(1+\phi^2)^2}x_t(x_{t-1}+x_{t+1})^3 + \frac{\phi^2}{(1+\phi^2)^4}(x_{t-1}+x_{t+1})^4\right]$$

$$= \left[3\phi^2 - 12\phi^2 + 2\left[\frac{2\phi^2}{(1+\phi^2)^2}+1\right](1+5\phi^2)\right.$$

$$\left. -24\frac{\phi^2}{(1+\phi^2)^2}(1+\phi^2) + \frac{12\phi^2(1+\phi^2)^2}{(1+\phi^2)^4}\right]\frac{\sigma^4}{(1-\phi^2)^2}$$

$$= \left[(3\phi^2-12\phi^2+2+10\phi^2) + \frac{1}{(1+\phi^2)^2}(4\phi^2-24\phi^2+12\phi^2+20\phi^4-24\phi^4)\right]\frac{\sigma^4}{(1-\phi^2)^2}$$

$$= [(2+\phi^2)(1+\phi^2)^2 - 8\phi^2 - 4\phi^4]\frac{\sigma^4}{(1+\phi^2)^2(1-\phi^2)^2}$$

$$= (2-3\phi^2+\phi^6)\frac{\sigma^4}{(1+\phi^2)^2(1-\phi^2)^2} \tag{B.11}$$

Since $E\,\Delta_{t_0}(t_0;\phi) = 0$,

$$E\left[-\phi x_t^2 + x_t(x_{t-1}+x_{t+1}) - \frac{\phi}{(1+\phi^2)}(x_{t-1}+x_{t+1})^2\right] = -\frac{\sigma^2\phi}{(1+\phi^2)} \quad . \tag{B.12}$$

Thus, squaring (4.10) and taking expectations, (B.11) and (B.12) yield

$$C_\phi = \left[\frac{\phi}{(1+\phi^2)}\right]^2 + 2\left[\frac{\phi}{(1+\phi^2)}\right]\left[-\frac{\phi}{(1+\phi^2)}\right] + \frac{2-3\phi^2+\phi^6}{(1-\phi^2)^2(1+\phi^2)^2}$$

$$= (-\phi^2(1-\phi^2)^2 + 2 - 3\phi^2 + \phi^6)\,\frac{1}{(1-\phi^2)^2(1+\phi^2)^2}$$

$$= \frac{2}{(1+\phi^2)^2} \tag{B.13}$$

APPENDIX C: Computation of EDV

The computations for EDV parallel that for EDC (see Appendix C), though we must compute the score function for σ^2. By (4.7), we need only compute $\Delta(t_0;\sigma^2)=\Psi_n^{(t_0)}(\sigma^2)-\Psi_n(\sigma^2)$, which is the difference between the score function for σ^2 with t_0 considered as missing and the score function with all the data for a perfectly observed Gaussian process.

The score function with no missing data is

$$\Psi(\sigma^2) \;=\; -\frac{n}{2}\,\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{n} v_t^2/f_t \tag{C.1}$$

If x_{t_0} is missing, then the prediction errors and the variances are given by v_t^* and f_t^* in (B.3). Hence, for the leave-one-out case

$$\Delta(t_0;\sigma^2) \;=\; \frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}\left[\, v_{t_0+1}^{*2}/f_{t_0+1}^{*} - (v_{t_0+1}^2/f_{t_0+1}+v_{t_0}^2/f_{t_0})\right] \tag{C.2}$$

$$= \frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}\left[\frac{(x_{t_0+1}-\phi^2 x_{t_0-1})^2}{(1+\phi^2)} - \left((x_{t_0+1}-\phi x_{t_0})^2+(x_{t_0}-\phi x_{t_0-1})^2\right)\right]$$

$$= \frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}\Bigg[-(1+\phi^2)x_{t_0}^2+2\phi x_{t_0}(x_{t_0-1}+x_{t_0+1})$$

$$-\frac{2\phi^2}{(1+\phi^2)}x_{t_0-1}x_{t_0+1}+\left[\frac{1}{(1+\phi^2)}-1\right]x_{t_0+1}^2$$

$$+\left[\frac{\phi^4}{(1+\phi^2)}-\phi^2\right]x_{t_0-1}^2\Bigg]$$

$$= \frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}\Bigg[-(1+\phi^2)x_{t_0}^2+2\,\phi\, x_{t_0}(x_{t_0-1}+x_{t_0+1})$$

$$-\frac{\phi^2}{(1+\phi^2)}(x_{t_0-1}+x_{t_0+1})^2\Bigg]$$

which is the same as (4.15).

Assume $\mathbf{x}_t$ is a perfectly observed Gaussian process, and $\mathbf{y}_t$ behaves according to some outlier model. Proceeding as in Appendix C, define $D_{\sigma^2}=E(\Delta(t_0;\sigma^2)^2)$ and

$\delta_{y_t}(t_0;\sigma^2)=\Delta_{y_t}(t_0;\sigma^2)-\Delta(t_0;\sigma^2)$ where $\Delta_{y_t}(t_0;\sigma^2)$ is the difference in the score functions for y_t. Then

$$E(\Delta_{y_t}(t_0;\sigma^2)^2) = D_{\sigma^2}+E\,\delta_{y_t}(t_0;\sigma^2)^2 \tag{C.3}$$

As shown below, $D_{\sigma^2}=\dfrac{1}{2\sigma^4}$, and $E(\delta_{y_t}(t_0;\sigma^2)^2)$ is evaluated case by case below. From (C.3) we can compute EDV via (4.9).

AO Models:

We first consider an AO type outlier of size ζ at t_0. We can compute $\delta_{y_t}(t_0;\sigma^2)^2$ in the same fashion as Appendix B: replace x_{t_0} by $x_{t_0}+\zeta$ in (C.2) and subtract off $\Delta(t_0;\sigma^2)$. Using the same notation as before,

$$E\,\delta^{AO}_{(\zeta;t_0)}(t_0;\sigma^2)^2 = \frac{1}{4\sigma^8}E\left[\zeta^2(-(1+\phi^2))+\zeta(-2(1+\phi^2)x_{t_0}+2\phi(x_{t_0-1}+x_{t_0+1}))\right]^2$$

$$= \left[\frac{\zeta}{\sigma}\right]^4\frac{1}{4\sigma^4}(1+\phi^2)^2+\left[\frac{\zeta}{\sigma}\right]^2\frac{1}{4\sigma^4}\,\frac{1}{(1-\phi^2)}(4(1+\phi^2)^2-16\phi^2(1+\phi^2)+8\phi^2(1+\phi^2))$$

$$= \left[\frac{\zeta}{\sigma}\right]^4\frac{1}{4\sigma^4}(1+\phi^2)^2+\left[\frac{\zeta}{\sigma}\right]^2\frac{1}{\sigma^4}(1+\phi^2) \tag{C.4}$$

Adding D_{σ^2} to (C.4) and scaling by $2\sigma^4$ gives (4.16).

The calculations for when t_0-1 (or t_0+1) we left out proceed as follows:

$$E\,\delta^{AO}_{(\zeta;t_0)}(t_0+1;\sigma^2)^2 = \frac{1}{4\sigma^8}E\left[\zeta^2\left[-\frac{\phi^2}{(1+\phi^2)}\right]+\zeta\left[2\phi x_{t_0}-2\frac{\phi^2}{(1+\phi^2)}(x_{t_0-1}+x_{t_0+1})\right]\right]^2$$

$$= \left[\frac{\zeta}{\sigma}\right]^4\frac{1}{4\sigma^4}\,\frac{\phi^4}{(1+\phi^2)^2}+\left[\frac{\zeta}{\sigma}\right]^2\frac{1}{4\sigma^4}\,\frac{1}{(1-\phi^2)}\left[4\phi^2-16\frac{\phi^4}{(1+\phi^2)}+8\frac{\phi^4}{(1+\phi^2)}\right]$$

$$= \left[\frac{\zeta}{\sigma}\right]^4\frac{1}{4\sigma^4}\,\frac{\phi^4}{(1+\phi^2)^2}+\left[\frac{\zeta}{\sigma}\right]^2\frac{1}{\sigma^4}\,\frac{\phi^2}{(1+\phi^2)} \tag{C.5}$$

This leads to (4.17).

For outliers of size ζ at $t_0 - 1$ and $t_0 + 1$, it is easy to see (as in (C.4)) that $E\,\delta^{AO}_{(\zeta; t_0-1, t_0+1)}(t_0; \sigma^2) = E\,\delta^{AO}_{(2\zeta; t_0)}(t_0 - 1; \sigma^2)$, which is given by (C.5).

IO Models:

Proceeding in the usual way, (4.22) is derived from

$$E\,\delta^{IO}_{(\zeta, t_0)}(t_0; \sigma^2)^2 = \frac{1}{4\sigma^8} E\left[\zeta^2\left[-(1+\phi^2)+2\phi^2-\frac{\phi^4}{1+\phi^2}\right]\right.$$

$$\left. + \zeta\left[(-2(1+\phi^2)+2\phi^2)\,x_{t_0} + \left[2\phi - \frac{2\phi^3}{(1+\phi^2)}\right](x_{t_0-1}+x_{t_0+1})\right]\right]^2$$

$$= \left[\frac{\zeta}{\sigma}\right]^4 \frac{1}{4\sigma^4}\left[(\phi^2-1)-\frac{\phi^4}{(1+\phi^2)}\right]^2 + \left[\frac{\zeta}{\sigma}\right]^2 \frac{1}{4\sigma^4}\,\frac{1}{(1-\phi^2)}\left[4-\frac{16\phi^2}{1+\phi^2}+8\,\frac{\phi^2}{1+\phi^2}\right]$$

$$= \left[\frac{\zeta}{\sigma}\right]^4 \frac{1}{4\sigma^4}\,\frac{1}{(1+\phi^2)^2} + \left[\frac{\zeta}{\sigma}\right]^2 \frac{1}{\sigma^4}\,\frac{1}{(1+\phi^2)}$$

Also, $EDV^{IO}_{(\zeta; t_0)}(t; \sigma^2) = 2\sigma^4 D_{\sigma^2}$ for $t > t_0$ since

$$E\,\delta^{IO}_{(\zeta, t_0)}(t_0+i; \sigma^2)^2 = \frac{1}{4\sigma^8} E\left[\zeta^2(-(1+\phi^2)\phi^{2i}+2(1+\phi^2)\phi^{2i}-(1+\phi^2)\phi^{2i})\right.$$

$$\left. + \zeta(-2(1+\phi^2)\phi^i + 2(1+\phi^2)\phi^i)\,x_{t_0} + (2\phi^{i+1}-2\phi^{i+1})(x_{t_0-1}+x_{t_0+1})\right]^2$$

$$= 0$$

Computation of D_{σ^2}

Straightforward algebra, using the expectations given in (B.10), shows that

$$E\left[-(1+\phi^2)x_t^2 + 2\phi x_t(x_{t-1}+x_{t+1}) - \frac{\phi^2}{(1+\phi^2)}(x_{t-1}+x_{t+1})^2\right]^2$$

$$= E\left[(1+\phi^2)^2 x_t^4 - 4\phi(1+\phi^2)x_t^3(x_{t-1}+x_{t+1})\right.$$

$$+ 6\phi^2 x_t^2 (x_{t-1} + x_{t+1})^2$$

$$\left. - \frac{4\phi^3}{(1+\phi^2)} x_t (x_{t-1} + x_{t+1})^3 + \frac{\phi^4}{(1+\phi^2)^2} (x_{t-1} + x_{t+1})^4 \right]$$

$$= \Big[3(1+\phi^2)^2 - 24\phi^2(1+\phi^2) + 12\phi^2(1+5\phi^2)$$

$$\left. - 48\phi^4 + 12\phi^4 \right] \frac{\sigma^4}{(1-\phi^2)^2}$$

$$= [3 - 6\phi^2 + 3\phi^4] \frac{\sigma^4}{(1-\phi^2)^2}$$

$$= 3\sigma^4$$

Since $E\,\Delta_{y_t}(t;\sigma^2) = 0$, then

$$E\left[-(1+\phi^2)x_t^2 + 2\phi x_t (x_{t-1} + x_{t+1}) - \frac{\phi^2}{(1+\phi^2)} (x_{t-1} + x_{t+1})^2 \right] = -\sigma^2$$

Hence,

$$D_{\sigma^2} = \frac{1}{4\sigma^4} + 2\left[\frac{1}{2\sigma^2}\right]\left[-\frac{1}{2\sigma^2}\right] + \left[\frac{1}{2\sigma^4}\right]^2 (3\sigma^4)$$

$$= \frac{1}{2\sigma^4}\ .$$

References

Atkinson, A.C. (1985). *Plots, Transformations, and Regressions: An Introduction to Graphical Methods of Diagnostic Regression Analysis.* Oxford: Clarendon Press.

Becker, R.A., and Chambers, J.M. (1984). *S: An interactive Environment for Data Analysis and Graphics.* Belmont, California: Wadsworth Advanced Book Program.

Belsley, D.A., Kuh, E., and Welsch, R.E. (1980). *Regression Diagnostics.* New York: Wiley.

Box, G.E.P., and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control.* Oakland, CA: Holden-Day.

Box, G.E.P., and Tiao, G.C. Intervention analysis with applications to economic and enviromental problems. *Journal of the American Statistical Association,* **70,** *70-79.*

Brillinger, D.R. (1966). Discussion of Mr. Sprent's apaper. *Journal of the Royal Statistical Society,* Series B, **28,** 294.

Burman, J.P. (1985). Report on ASA Fellowship 1984–5. Statistical Research Division, U.S. Bureau of the Census, Room 3524, FOB #3, Washington, D.C. 20233.

Cook, R.D., and Weisberg, S. (1982). *Residuals and Influence in Regression.* New York: Chapman and Hall.

Denby, L., and Martin, R.D. (1979). Robust estimation of the first-order autoregressive parameter. *Journal of the American Statistical Association,* **74,** 140–146.

Durbin, J. (1960). The fitting of time series models. *Review of the International Institute of Statistics,* **28,** 233–244.

Efron, B., and Hinckley, D.V. (1978). Assessing the accuracy of the maximum likelihood estimation: observed versus expected Fisher information. *Biometrika,* **65,** 457–487.

Engle, R.F. and Kraft, D.F. (1983) *Applied Time Series Analysis of Economic Data,* ed. A. Zellner. Washington, D.C.: U.S. Bureau of Census, 176-177.

Fox, A.J. (1972). Outliers in time series. *Journal of the Royal Statistical Society,* Ser. B, **34,** 350–363.

Fuller, W.A. (1976). *Introduction to Statistical Time Series.* New York: John Wiley.

Gardner, G., Harvey, A.C., Phillips, G.D.A. (1980). An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. *Applied Statistics,* **29,** 311–322.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society, Ser. B,* **38,** *205-247.*

Harvey, A.C. (1981). *Time Series Models*. John Wiley: New York.

Harvey, A.C., and Phillips, G.D.A. (1979)., Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika*, **66**, 49–58.

Harvey, A.C., and Pierse, R.G. (1984). Estimating missing observations in economic time series. *Journal of the American Statistical Association*, **7a**, 125–131.

Hillmer, S.C., Bell, W.R., and Tiao, G.C. (1983). Modeling considerations in the seasonal adjustment of economic time series. *Applied Time Series Analysis of Economic Data*. ed. A. Zellner. Washington, D.C.: U.S. Bureau of Census, 74-100.

Jones, R.H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics*, **22**, 389–395.

Kohn, R., and Ansley, C.F. (1986). Estimation, prediction, and interpolation for ARIMA models with missing data. *Journal of the American Statistical Association*, **81**, 751–761.

Levinson, N. (1947). The Wiener RMS error criterion in filter design and prediction. *Journal of Mathematical Physics*, **25**, 261–278.

Martin, R.D. (1979). Approximate conditional-mean type smoothers and interpolators. *Smoothing Techniques for Curve Estimation*, T. Basser and M. Rosenblatt, eds., 147-176, New York: Academic Press.

Martin, R.D. (1981). Robust methods for time series. *Directions in Time Series*, D.F. Findley, ed., 683-759, New York: Academic Press.

Martin, R.D., Samarov, A. and Vandaele, W. (1983). Robust methods for ARIMA models. *Applied Time Series Analysis of Economic Data*, ed. A. Zellner. Washington, D.C.: U.S. Bureau of Census, 153-169.

Martin, R.D. and Su, K.Y. (1985). Robust filters and smoothers: definition and design. *Technical Report No. 58*, Department of Statistics, University of Washington, Seattle, WA, 98195.

Martin, R.D., and Thompson, D.J. (1982). Robust-resistant spectrum estimation. *Proceedings of the IEEE* **70**, 1097-1115.

Martin, R.D., and Zeh, J.E. (1977). Determining the character of time series outliers. *1977 Business and Economic Statistics* Section *Proceedings of the American Statistical Association*, 818–823.

Martin, R.D., and Yohai, V.J. (1986). Influence curves for time series. *Annals of Statistics*, **11**, 781–818.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, **9**, 705–724.

Schweppe, F. C. (1973). *Uncertain Dynamic Systems*, Prentice-Hall.

Smith, A.F.M., and West,M. (1983). **Monitoring renal** transplants: an application of the **Multiprocess Kalman filter.** *Biometrics* **39,** *867-878*.

Storer, B.E., and Crowley, J. (1985). A **diagnostic** for Cox regression and generalized **conditional likelihoods.** *Journal of the American Statistical Association*, **80,** 139–147.

Tsay, R.S. (1986). **Time series model specification in the** presence of outliers. *Journal of the American statistical Association*, **81,** 132–141.

Weisberg, S. (1980). *Applied Linear Regression*. John Wiley: New York.

West, M. (1986). **Bayesian model monitoring.** *Journal of the Royal Statistical Society, Ser. B*, **48,** *70-78*.

West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, **80** *73-98*.

END
DATE
FILMED
5-88
DTIC